



# Modélisation, purification et caractérisation des modules et domaines de la PI4KA humaine

Cyntia Taveneau

## ► To cite this version:

Cyntia Taveneau. Modélisation, purification et caractérisation des modules et domaines de la PI4KA humaine. Biochimie [q-bio.BM]. Université Paris Sud - Paris XI, 2015. Français. NNT : 2015PA114827 . tel-01362510

**HAL Id: tel-01362510**

**<https://theses.hal.science/tel-01362510>**

Submitted on 9 Sep 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Comprendre le monde,  
construire l'avenir®



UNIVERSITÉ PARIS-SUD

ÉCOLE DOCTORALE 425 :  
INNOVATION THÉRAPEUTIQUE : DU FONDAMENTAL À L'APPLIQUÉ

Laboratoire : Institut de Biologie Intégrative de la Cellule  
*Département de Biochimie, Biophysique et Biologie Structurale &  
Département de Virologie*

## THÈSE DE DOCTORAT

BIOCHIMIE ET BIOLOGIE STRUCTURALE  
par

**Cyntia TAVENEAU**

Modélisation, purification et caractérisation des modules et  
domaines de la PI4KA humaine

Date de soutenance : 08/09/2015

**Composition du jury :**

Directeur de thèse :	Stéphane BRESSANELLI	DR2 (I2BC, département B3S)
Rapporteurs :	François PENIN Xavier HANOULLE	IR (IBPC UMR 5086 CNRS & Université de Lyon) CR (INSB UMR 8576 CNRS & Université de Lille)
Examineurs :	Catherine ETCHEBEST Herman VAN TILBEURGH	Professeur (DSIMB UMR_S 1134 INSERM & Université Paris Diderot) Professeur (I2BC, département B3S & Université Paris-Sud XI)



# Résumé

---

La phosphatidylinositol-4-kinase de type III $\alpha$  est une kinase de lipide eukaryote ubiquitaire qui synthétise le phosphatidylinositol-4-phosphate PtdIns(4)P de la membrane plasmique. Ce phosphoinositide est d'autant plus important qu'il tient un rôle majeur dans différentes voies de signalisation cellulaire, le trafic vésiculaire ainsi que dans l'identité des organelles. De plus, la PIK4A humaine est un facteur essentiel pour la réplication du virus de l'hépatite C (VHC). En effet, le recrutement du complexe de réplication du VHC par la protéine virale NS5A à la membrane du reticulum endoplasmique permet la formation d'un réseau membranaire à l'origine de la structuration des complexes de replication viraux.

La PI4KA est une kinase imposante (2102 résidus, 240 kDa pour la PI4KA humaine) qui possède un domaine kinase C-terminal d'environ 400 résidus précédé d'un domaine formé de répétitions Armadillo pour lequel aucune fonction n'a été déterminée. Le rôle ainsi que le repliement des 1500 résidus N-terminaux de PI4KA ne sont pas connus à ce jour.

Afin d'en savoir plus sur la structure tri-dimensionnelle de la PI4KA humaine, nous avons utilisé des outils bio-informatiques afin de délimiter et de modéliser les modules et domaines la composant. Nous avons pu ainsi les exprimer et les produire en bactérie et en cellules d'insecte afin de vérifier nos hypothèses. Nous avons pu conclure que PI4KA est composée de deux modules. Le module N-terminal (1100 résidus), est composé de deux domaines dont un solénoïde  $\alpha$ . Les résultats obtenus par diffusion des rayons X aux petits angles (SAXS) nous permettent de définir leur agencement potentiel.

Le second module (1000 résidus), le module C-terminal, est l'enzyme-core. Son analyse nous a permis d'identifier une similitude remarquable avec les sérine/thréonine kinases, comme mTor, apparentées aux phosphatidylinositol-3-kinases. Nous avons défini au début du module C-terminal de PI4KA trois domaines putatifs que nous avons nommés DI, DII et DIII. Nos collaborateurs ont montré qu'ils sont essentiels à l'activité kinase de la protéine ainsi qu'à la réplication du VHC. Le domaine DI a été caractérisé et a permis la validation d'une nouvelle paramétrisation de la molécule de N, N-diméthyl-dodécylamine oxide (LDAO) pour des simulations de dynamique moléculaire. Enfin, la PI4KA humaine dans son entier a été exprimée en cellules d'insecte puis purifiée, et un premier test d'interaction avec les membranes a été initié.

## Mots clés

PI4KA; Bio-informatique, Modélisation; Expression ; Purification; LDAO; SAXS, NS5A





# Remerciements

---

Je tiens tout d'abord à remercier François Penin ainsi que Xavier Hanoulle de me faire l'honneur d'être les rapporteurs de mes travaux de thèse. Je tiens également à remercier Herman Van Tilbeurgh et Catherine Etchebest de participer à mon jury de thèse.

Je tiens à exprimer ma gratitude à mon directeur de thèse, Stéphane Bressanelli, qui a cru en moi dès mon stage de deuxième année de master. Je tiens particulièrement à le remercier pour sa confiance, le degré de liberté qu'il m'a laissé quant à la conduite de mes travaux ainsi que pour sa disponibilité. Sa rigueur scientifique sera toujours un fil conducteur pour moi. Merci encore pour ton optimisme inconditionnel et ton soutien !

Je tiens aussi à remercier l'Agence Nationale de Recherche sur le SIDA et les Hépatites virales (ANRS) qui a financé trois ans de cette thèse.

Je tiens à remercier tous les membres du laboratoire de Virologie Moléculaire et Structurale du CNRS de Gif-sur-Yvette dans lequel j'ai réalisé la majeure partie de ma thèse. Tout naturellement, je tiens à remercier Yves Gaudin de m'avoir permis de mener à bien ce projet. Je remercie plus particulièrement Isabelle Auzat de m'avoir permis d'utiliser son travail portant sur NS5A afin de pouvoir exprimer et purifier le domaine d1 de la protéine virale au laboratoire. Un merci à Nathalie qui m'a aidé à prendre mes marques dans le laboratoire et sans qui le café aurait été moins girly. Un merci tout particulier à Matthieu pour m'avoir écouté quand j'avais besoin « de faire le 'ti mec » et sans qui mes révélations de western blot auraient été beaucoup plus courtes. Merci aussi à Momolière pour ses débats scientifiques en tout genre et Nolwenn pour sa bonne humeur et son humour. Malheureusement je ne peux pas m'étaler sur la moitié du manuscrit, et je vais donc m'arrêter là... mais un 1/5 ème de page ne suffit pas à exprimer tout ce que j'ai pu vivre en quatre ans au laboratoire !

Je remercie également Karine Blondeau, maître de conférences à l'université Paris-Sud XI, pour son expertise sur la sur-expression des protéines en *E.Coli* cultivées en fermenteur. Merci beaucoup pour ta grande implication dans le projet ainsi que pour ta disponibilité, ta pédagogie et surtout ta bonne humeur !

Je tiens à remercier Maïté sans qui j'aurais écrit mon manuscrit dans des conditions bien plus difficiles ainsi que pour la pêche qu'elle me donne le matin (et pas qu'à cause du café). Reste comme tu es, tu es géniale !! Merci à Thibault et Pierre qui m'ont appris à me poser les bonnes questions... « Est-ce qu'on peut s'en servir pour donner de l'élan à un pigeon ? » et autres questionnements existentiels. Merci à Sonia et Yves pour leur accueil, leur jovialité, leur aide ainsi que leur soutien.

Un grand merci à ma famille pour son soutien moral, tout particulièrement mes parents et mon (pas si) petit frère. Votre soutien a été sans borne, et je sais que vous serez toujours là quoique je choisisse pour la suite. Merci aussi pour l'intérêt que vous avez porté au déroulement même de ma thèse.

On dit que les chercheurs finissent par personnaliser leur protéine. Vous êtes donc des graines de génie en devenir (je ne sais d'ailleurs toujours pas si ma « protéinine est vivante » ou pas). Merci aussi fréro et d'ailleurs bon courage à toi maintenant !

Un grand merci à Amandine pour les pique-niques, les cinés, les voyages, les sorties nocturnes.... Notre to-do liste est encore bien longue, nous avons encore des années prometteuses devant nous ! Qui sait, peut-être que le prochain voyage se fera à la Réunion. Encore merci de m'avoir ouvert les yeux il y a bien longtemps de cela maintenant ! Et évidemment merci à Vincent, Micka, Estelle, Angélique, toujours présents même dans les moments les plus difficiles.

Enfin merci à Sofiane, sans qui j'aurais certainement mangé des coquillettes pendant toute la rédaction. Merci pour ton soutien inconditionnel, ton accompagnement, ta folie, ta spontanéité et pour tout ce qui fait de notre quotidien un plaisir de tous les jours.

# Publications & Travaux réalisés pendant cette thèse

---

**SEUL CE PROJET FAIT L'OBJET DU MANUSCRIT**

**Publications**

- *Mapping of functional domains of the lipid kinase phosphatidylinositol 4-kinase type III alpha involved in enzymatic activity and hepatitis C virus replication.*

Harak C, Radujkovic D, Taveneau C, Reiss S, Klein R, Bressanelli S, Lohmann V.  
J Virol. 2014 Sep 1;88(17):9909-26. doi: 10.1128/JVI.01063-14. Epub 2014 Jun 11.

- *Definition and expression in E. coli of large fragments from the human lipid kinase Phosphatidylinositol 4-Kinase Type III Alpha, and purification of a 1100-residue N-terminal module*

Cyntia Taveneau, Karine Blondeau, Stéphane Bressanelli  
Protein Expr Purif. 2015 Jul 2;114:121-127

- *A new potential for N,N-Dimethyldodecylamine-N-oxide surfactant based on the Amber force field for Molecular Dynamics Simulations of Micelles and application to a fragment of the lipid kinase PI4KA*

Esra Karakas \*, Cyntia Taveneau\*, Stéphane Bressanelli, Massimo Marchi, Bruno Robert and Stéphane Abel ; **en cours de réécriture**

*\*co-premières auteurs*

**Présentations orales**

2015 : 15<sup>ème</sup> Réunion du Réseau National Hépatites  
GdR 3696 protéines membranaires, aspects moléculaires et cellulaires

2012 : 12<sup>ème</sup> Réunion du Réseau National Hépatites  
GdR 3334 Supra-MemBio « assemblage supramoléculaires et membranes biologiques »

**Posters**

2014 : XVI<sup>ème</sup> Journées Francophones de Virologie

2013: Formation “Méthodes avancées en biologie structurale intégrée de protéines membranaires”

2012: XIV<sup>ème</sup> Journées Francophones de Virologie

## **Compétences**

Biologie moléculaire

*Clonage*

Biochimie

*Expression de protéines en cellules d'insecte et bactérie, Purification de protéines*

Bio-informatique

*Modélisation par homologie et par threading, Prédiction de structure, Alignement de séquence*

Biophysique

*SEC-MALLS, Dichroïsme Circulaire, Fluorescence des tryptophanes, SAXS*

## **Encadrement**

2013 (2 mois) : étudiante en 2<sup>ème</sup> année de licence BioPlus (Université Paris-Sud XI)

*Expression et purification du domaine DI de PI4KA*

2014 (2 mois) : étudiant en 2<sup>ème</sup> année de BTS Bio-analyse et Contrôles (Ecole Technique Supérieure de Laboratoire)

*Mise au point de l'expression et de la purification du module N-terminal de PI4KA*

---

### *Projet NS5-polymérase de la dengue*

---

#### **DETERMINATION DE L'ORIENTATION RELATIVE DES DEUX DOMAINES DE LA POLYMERASE DU VIRUS DE LA DENGUE DE TYPE II PAR SAXS**

## **Posters**

2012: Réunion du Groupe Thématique Biologie de l'Association Française de Cristallographie (GTBio)

2014: Journée de l'Ecole doctorale 425 « Innovation Thérapeutique, du fondamental à l'appliqué »

## **Compétences**

Biophysique et Bio-informatique

*SEC-MALLS, SEC-SAXS (collecte et traitement de données, modélisation de la structure de NS5 entière par combinaison de structure cristallographique et de données basse résolution)*

---

## *Projet Septine 9 -Collaboration*

---

### **OPTIMISATION DE L'EXPRESSION EN BACTERIE ET DE LA PURIFICATION A GRANDE ECHELLE DE LA SEPTINE 9 ET DE TROIS DE SES MUTANTS**

#### **Publications**

*Hepatitis C Virus promotes the accumulation of lipid droplets through a Septin 9-PtdIns5P-ERK cascade*

Akil Abdallah<sup>1</sup>, Juan Peng, Mohyeddine Omrane, Claire Gondeau, Mickaël Marin, Hélène Tronchère, Cyntia Taveneau, Célia-Caillet-Saguy, Sokhavuth Sar, Philippe Briolotti, Soumaya Benjelloun, Abdelaziz Benjouad, Patrick Maurel, Valérie Thiers, Stéphane Bressanelli, Didier Samuel, Christian Bréchet and Ama Gassama-Diagne ; **Soumis (résumé Annexe 12)**

#### **Encadrement**

2013 (2 mois) : étudiant en 2<sup>ème</sup> année de BTS Bio-analyse et Contrôles (Ecole Technique Supérieure de Laboratoire)

*Optimisation de l'expression et de la purification de la Septine 9 humaine et de trois de ses mutants*

---

## *Projet PrP -Collaboration*

---

### **CRISTALLISATION ET RESOLUTION DE LA STRUCTURE DE MUTANTS DE PRION (HUMAN REZAEI ET DAVY MARTIN, INRA)**

#### **Compétences**

Biophysique

*Cristallographie des rayons X, résolution de structure par remplacement moléculaire*

# SOMMAIRE

<b>RESUME .....</b>	<b>2</b>
<b>REMERCIEMENTS .....</b>	<b>4</b>
<b>PUBLICATIONS &amp; TRAVAUX .....</b>	<b>6</b>
<b>ABREVIATIONS.....</b>	<b>12</b>
<b>INTRODUCTION .....</b>	<b>16</b>
<b>I. LE VIRUS DE L'HÉPATITE C .....</b>	<b>18</b>
<b>A. Hépatite C : problème de santé publique mondial .....</b>	<b>18</b>
1) Avancées majeures de la recherche sur le VHC.....	19
2) Mode de transmission .....	19
3) Histoire naturelle.....	20
4) Cas particulier des co-infections.....	20
5) Traitement.....	21
<b>B. Hétérogénéité génétique du VHC .....</b>	<b>22</b>
1) CLASSIFICATION .....	22
2) Hétérogénéité géographique du virus.....	23
3) Grande variabilité génétique du virus .....	23
<b>C. Le virus de l'hépatite C.....</b>	<b>24</b>
1) STRUCTURE DU VIRUS DE L'HÉPATITE C.....	24
2) Organisation du génome viral .....	25
3) Cycle viral du VHC.....	27
<b>D. Réseau membranaire.....</b>	<b>28</b>
1) Un mécanisme utilisé chez les virus à ARN positif .....	28
2) MORPHOLOGIE DU RÉSEAU MEMBRANAIRE CHEZ LE VHC.....	29
3) Protéines de l'hôte impliquées dans la formation du réseau membranaire .....	30
4) PI4KA : une protéine cellulaire essentielle à la réplication du VHC .....	31
<b>II. LA PHOSPHATIDYLINOSITOL-4-KINASE IIIA HUMAINE (PI4KA) .....</b>	<b>33</b>
<b>A. Les phosphoinositides : lipides essentiels à la cellule.....</b>	<b>33</b>
1) IMPORTANCES DES PHOSPHOINOSITIDES .....	33
2) Nature chimique des phosphoinositides .....	34
3) Métabolisme des phosphoinositides.....	35
4) Le phosphatidylinositol(4)phosphate .....	37
<b>B. La phosphatidylinositol-4-kinase de type IIIα PI4KA.....</b>	<b>39</b>
1) LES PI4-KINASES.....	39
2) Rôle cellulaire de PI4KA.....	40
3) Localisation et recrutement de PI4KA .....	40
4) Éléments structuraux connus de PI4KA et de ses homologues .....	42
i) Domaine Catalytique.....	42
ii) Domaine à répétitions Armadillo et modules d'interaction.....	43
iii) Cas particulier des Phosphatidylinositol-3 kinase-related kinases (PIKKs) .....	43
5) Interaction PI4KA-NS5A.....	45
i) Rôle et Structure de NS5A.....	45
ii) Interaction entre NS5A et PI4KA humaine .....	46
6) Interaction PI4KA-NS5B .....	47
7) PI4KA, cible potentielle pour un antiviral anti-VHC ? .....	48



## **PARTIE I : DEFINITION DES MODULES ET DOMAINES COMPOSANT LA PI4KA HUMAINE ET FONCTION**

### **POTENTIELLE.....50**

#### **I. DÉFINITION DES DOMAINES C-TERMINAUX DE PI4KA PAR BIO-INFORMATIQUE : VERS UNE MEILLEURE COMPRÉHENSION DE LA RÉPLICATION DU VHC ET DE LA RÉGULATION DE L'ACTIVITÉ CATALYTIQUE DE PI4KA..... 52**

##### **A. Article n°1- Mapping of functional domain of the lipid kinase phosphatidylinositol 4-Kinase type III alpha involved in enzymatic activity and hepatitis c virus replication ..... 52**

1) CE QU'IL FAUT RETENIR.....	52
2) ARM-2102, étude détaillée et comparaison avec les homologues de structure connue .....	76
i) Domaine ARM .....	76
ii) Domaine catalytique .....	77
iii) ETUDE BIO-INFORMATIQUE DES DOMAINES PUTATIFS DI-DIII ET DU MODULE N-TERMINAL 1-1118 .....	81
3) Etude bio-informatique de la partie N-terminale de PI4KA et du module DI-DIII .....	81
i) Matériel et méthode .....	82
▪ Alignement multiple par ClustalΩ .....	82
▪ Recherche d'homologues lointains : HHblits, HHpred et Phyre2.....	82
▪ Analogues structuraux : modélisation par threading.....	83
ii) Résultats.....	83
▪ Deux modules potentiels : 1-1100 et DI-2102 .....	83
▪ Etude du module DI-DIII.....	84
▪ Etude du module N-terminal 1-1118 .....	86

##### **B. Délimitation expérimentale des modules et domaines composant PI4KA ..... 92**

1) ARTICLE N°2: DEFINITION, EXPRESSION IN E.COLI AND PURIFICATION OF LARGE FRAGMENTS FROM THE HUMAN LIPID KINASE PI4KA.....	92
2) Complément d'article : Expression et solubilité des constructions C-terminales.....	101
i) Matériel et méthode .....	101
▪ Mise au point de l'expression des protéines en E.Coli.....	101
▪ Tests de purification en batch.....	102
▪ Test de purification à grande échelle (constructions GST seulement).....	102
ii) Résultats.....	104
▪ Constructions C-terminales.....	104
▪ Constructions N-terminales .....	106
3) Mise au point de l'expression et de la purification de 26-1118 à grande échelle .....	107
i) Matériel et méthode .....	107
▪ Choix de la souche bactérienne et de la température d'induction .....	107
▪ Optimisation des conditions de culture : choix du milieu et culture en fermenteur .....	107
ii) Résultats.....	110
▪ Expression de GST 26-1118.....	110
▪ Purification de GST 26-1118.....	113

## **PARTIE II : EXPRESSION A GRANDE ECHELLE, PURIFICATION ET CARACTERISATION BIOPHYSIQUE DE DEUX**

### **COMPOSANTES DE PI4KA : MODULE N-TERMINAL ET DOMAINE PUTATIF DI..... 117**

#### **I. CARACTÉRISATION DU MODULE N-TERMINAL DE PI4KA 26-1118..... 119**

##### **A. Matériel et méthode..... 119**

1) EVALUATION DU RAYON DE STOKES PAR CHROMATOGRAPHIE D'EXCLUSION .....	119
2) Etude de la thermostabilité par fluorescence .....	119
3) Influence du sel sur la stabilité et la solubilité de.....	120
26-1118.....	120
4) Diffusion des rayons X aux petits angles (SAXS) .....	121
i) Principe <sup>122</sup> .....	121
ii) Acquisition des données SAXS.....	122
iii) Traitement de données.....	122
▪ Traitement des données brutes.....	123
▪ Construction de l'enveloppe de la protéine ab initio .....	126
5) Cristallogénèse .....	127

B.	Résultats .....	128
II.	CARACTÉRISATION DU DOMAINE PUTATIF DI .....	135
A.	Article n°3: A new potential for N,N-Dimethyldodecylamine-N-oxide surfactant based on the Amber force field for Molecular Dynamics Simulations of Membrane Proteins and application to a fragment of the lipid kinase PI4KA .....	135
1)	CARACTÉRISATIONS SUPPLÉMENTAIRES DE DI ET RÉSULTATS PRÉLIMINAIRES D'INTERACTION AVEC D1-NS5A.....	194
i)	Matériel et méthode .....	194
▪	Caractérisation : .....	194
▪	Interaction avec le domaine d1 de NS5A .....	195
ii)	Résultats.....	197
<b>PARTIE III : PROTEINE PERIPHERIQUE A LA MEMBRANE : SOLUBILITE DE PI4KA, INTERACTION A LA MEMBRANE ET MODELE COMPLET .....</b>		<b>202</b>
I.	MATÉRIEL ET MÉTHODE .....	204
A.	Expression en baculovirus de PI4KA entière et de 1129-2102 .....	204
1)	CONSTITUTION DU STOCK DE BACULOVIRUS PAR SYSTÈME BAC-TO-BAC <sup>138</sup> .....	204
i)	Constitution du bacmide en <i>E.Coli</i> DH10Bac.....	204
ii)	Constitution du stock de baculovirus .....	205
iii)	Test de Multiplicité d'Infection (MOI) .....	206
iv)	Choix de la souche de cellule d'insecte .....	207
B.	Test de solubilité et interaction avec les membranes.....	207
1)	TESTS DE SOLUBILITÉ DE PI4KA ENTIÈRE ET DE 1129-2102 .....	207
2)	Mise au point de la purification de PI4KA entière .....	208
3)	Flottaison de membrane .....	209
4)	Tests de protéolyses ménagées.....	210
II.	RÉSULTATS.....	211
III.	DISCUSSION : SIMILITUDE AVEC LES PIKKS, STRUCTURE COMPLÈTE DE PI4KA ET FIABILITE DE LA MODELISATION.....	217
<b>ANNEXES .....</b>		<b>225</b>
	Annexe 1 : alignement de séquence du domaine catalytique.....	227
	Annexe 2 : alignement de séquence module DI-DIII .....	231
	Annexe 3 : ALIGNEMENT de séquences du module N-terminal des pi4ka/STT4p .....	233
	ANNEXE 4 : expression en bactérie des constructions c-terminales et solubilité .....	236
	Annexe 5 : purification des constructions C-terminales.....	237
	Annexe 6 : conditions d'expression testées des constructions N-terminales en bactérie et solubilité ...	238
	Annexe 7 : Tests de purification des constructions N-terminales .....	239
	Annexe 8 : composition des différents milieux de culture utilisés pour la mise au point de l'expression de 26-1118 .....	240
	Annexe 9 : mise au point de la purification de 26-1118.....	241
	Annexe 10 : tests interaction entre DI de PI4KA et d1-NS5A .....	242
	Annexe 11. Protocole d'expression et de purification de LA PI4KA humaine entière mis au point par l'équipe de Volker Lohmann .....	243
	Annexe 12 : abstract de l'article ' Hepatitis C Virus promotes the accumulation of lipid droplets through a Septin 9-PtdIns5P-ERK cascade' .....	244
<b>BIBLIOGRAPHIE.....</b>		<b>246</b>



# Abréviations

---

Adaptator Binding Protein	ABD
Répétition de type Armadillo	ARM
Acide ribonucléique	ARN
Claudine 1	CLDN1
Concentration Micellaire Critique	CMC
Agents Anti-viraux Directs	DAA
Diacylglycérol	DAG
Diffusion Dynamique de la Lumière	DLS
Diamètre maximum	Dmax
Vésicule à double membrane	DMV
Focal Adhesion Targeting	FAT
<b>H</b> untingtin, <b>EF3</b> , <b>PP2A</b> , <b>TOR1</b>	HEAT
Modèle de Markov caché	HMM
HEAT related domain	HRD
Interféron $\alpha$ pégylé	IFN $\alpha$
Site d'entrée interne du ribosome	IRES
Lysogeny Broth	LB
Goutelette lipidique	LD
N, N-dimethyl-dodecylamine oxide	LDAO
Particule lipovirale	LVP
Milieu Auto-Inductible	MAI
Membrane plasmique	MP
Vésicule multimembranaire	MVB
Poids Moléculaire	MW
N-Octyl glucoside	NOG
Région non traduite	NTR
Non Structural	NS
Occludine	OCLN
Phase ouverte de lecture	ORF
Protéine Data Bank	PDB
Domaine homologue aux Pleckstrines	PH
Phosphoiositides	PI
Phosphatidylinositol kinase	PIK
Phosphatidylinositol-3 kinase-related kinases	PIKK
Phosphatidylinositol-4-kinase III $\alpha$	PI4KA
Phosphatidylinositol-X-phosphate	PtdIns(X)P
Fonction de paire	P(R)
Rat Sarcoma	Ras
Réticulum Endoplasmique	RE
Rayon de gyration	Rg
Rayon de Stokes	Rs

Diffusion des rayons X aux petits Angles	SAXS
Tetratricopeptide repeat	TPR
Three Discontinious Domain	TRD
Réseau trans-golgien	TGN
Virus de l'hépatite B	VHB
Virus de l'hépatite C	VHC
Virus de l'immunodéficience humaine	VIH

# Introduction

---



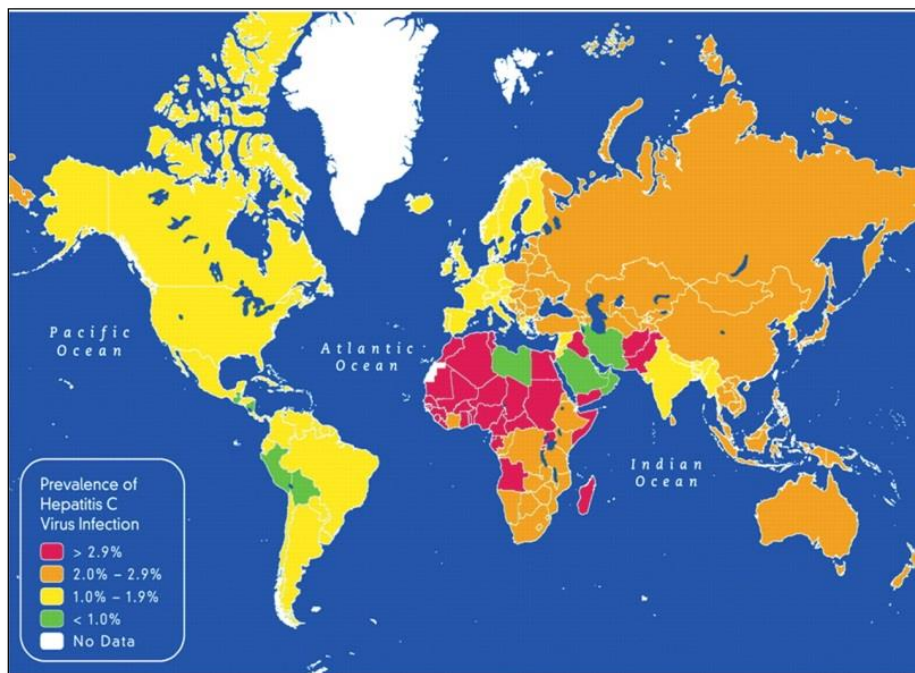
# I. Le virus de l'hépatite C

## A. Hépatite C : problème de santé publique mondial

Il peut être défini trois principales infections virales chroniques de par le monde : le SIDA causé par le VIH, l'hépatite B provoquée par le virus de l'hépatite B (VHB) et l'hépatite C.

L'hépatite C est une maladie du foie provoquée par un virus, le VHC, maladie présente dans le monde entier. L'OMS a évalué à plus de 350 000 par an le nombre de décès liés à des pathologies hépatiques provoquées par ce virus pour lequel il existe 130 à 150 millions de porteurs chroniques. Environ 15 à 45% des personnes infectées se débarrassent spontanément du virus dans les 6 mois tandis que pour les autres, l'infection évoluera vers la forme chronique de la maladie<sup>1</sup>.

La prévalence du VHC (figure 1) est plus élevée en Asie centrale et orientale, ainsi qu'en Afrique du Nord. Suite à une campagne d'éradication de bilharziose menée dans les années 50 dans des conditions insuffisamment stériles, l'Egypte est devenu le pays le plus touché par ce virus. Un intérêt tout particulier est depuis porté aux campagnes de prévention contre le VHC en Egypte, notamment en ce qui concerne les soins médicaux (transfusions sanguines, réutilisation d'aiguilles usagées)<sup>2</sup>.



**Figure 1-** Prévalence mondiale du virus de l'hépatite C chronique en 2012 <sup>3</sup>



## 1) Avancées majeures de la recherche sur le VHC

En 1975 a été mis en évidence un nouvel agent pathogène responsable d'hépatites autre que les hépatites A et B. Ces hépatites sont alors nommées non-A non-B<sup>2</sup>.

Le virus VHC lui-même sera identifié 15 ans plus tard, par Choo et al, dans le sérum d'un patient diagnostiqué hépatique non-A non-B<sup>29</sup>.

Cette découverte permit des avancées quant aux études *in vivo* de l'agent pathogène mais ne rendit pas possible les expériences *in vitro*. En effet, pendant plusieurs années, aucun modèle cellulaire n'a permis la bonne répllication du virus. La mise au point d'un système de réplicon dans des cellules Huh-7 HCC humaines en 1999 constitua alors une avancée majeure, permettant l'étude de la répllication intracellulaire du VHC, la découverte d'agents antiviraux et une meilleure compréhension des résistances aux médicaments. L'étude du cycle de répllication virale entier ne sera permise que suite à l'isolation de la souche d'hépatite C fulminante JFH1 en 2003, souche capable de se répliquer spontanément dans un certain nombre de lignées cellulaires<sup>4</sup>.

Les principales étapes marquantes de l'étude moléculaire et cellulaire du VHC sont répertoriées dans la chronologie ci-dessous

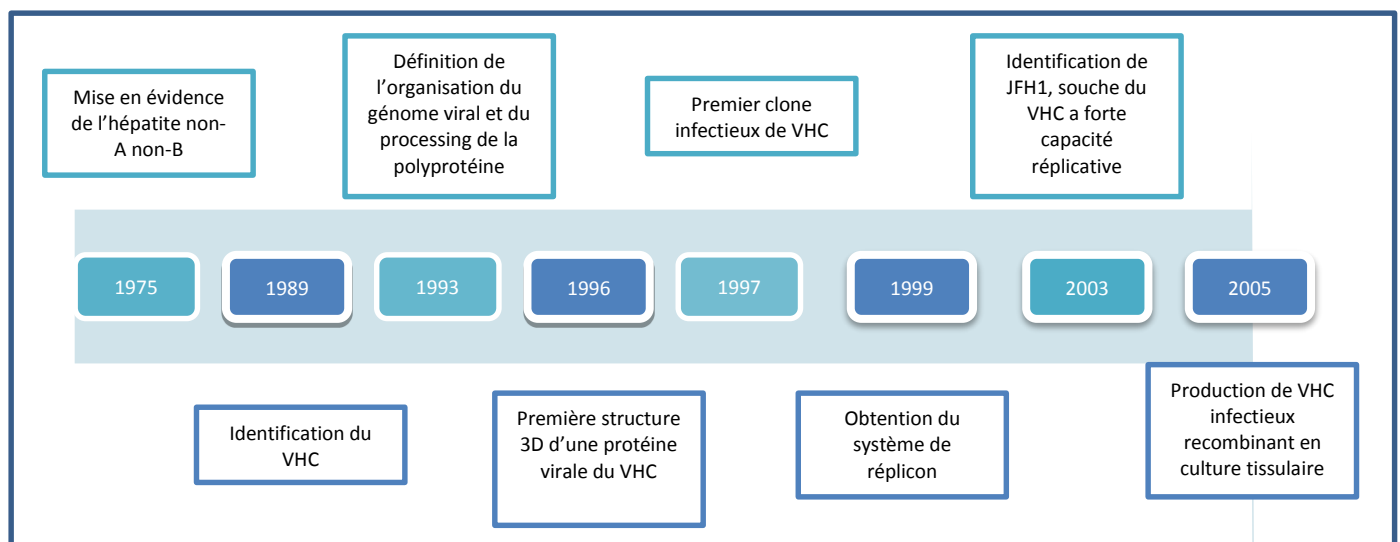


Figure 2 (Adapté de 4) - Chronologie des événements marquants la recherche sur le VHC.

## 2) Mode de transmission

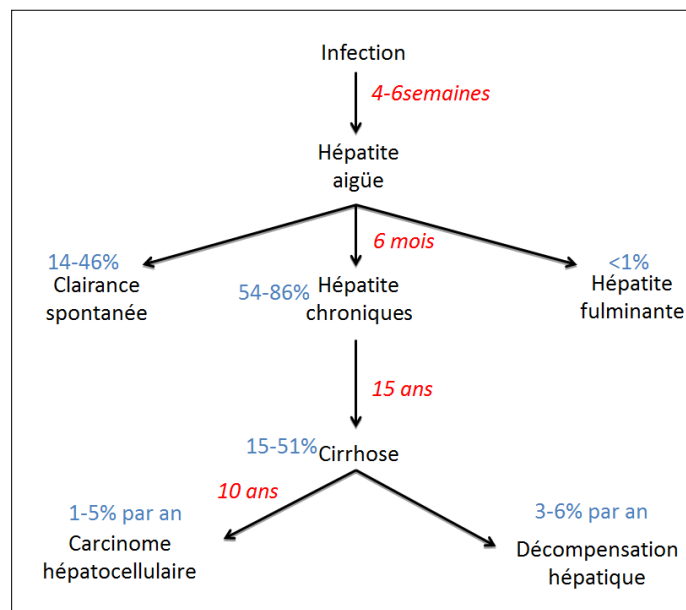
Le virus de l'hépatite C est un virus transmissible par contact sanguin direct. Le mode de transmission prépondérant du virus est fortement associé à une inégalité en matière de santé entre les pays. Ainsi, dans les pays peu développés ou en voie de développement, les infections par le VHC sont en grande partie dues aux procédures médicales insuffisamment stériles (transfusion sanguine, dialyse rénale) tandis que dans les pays développés, la première source d'infections provient de l'utilisation de seringues usagées pour injection de drogue ou bien l'utilisation de drogues elles-mêmes contaminées.

Le VHC peut aussi être transmis verticalement de la mère à l'enfant et par rapport sexuel, mais cela reste plus rare<sup>1, 5</sup>.

### 3) Pathogénèse de l'hépatite C

Après l'infection par le VHC, environ 80% des individus ne développent pas de symptômes, cependant la clairance du virus est observée dans une minorité de cas. 54% à 86% des patients adultes manifestent une hépatite dite chronique qui peut évoluer vers des fibroses, des cirrhoses ou encore des carcinomes hépatocellulaires. Un faible pourcentage des patients développe des manifestations extra-hépatiques.

L'évolution de l'infection par le VHC chronique varie suivant un certain nombre de facteurs extérieurs comme l'âge, les co-infections ou encore une grande consommation d'alcool <sup>1,6</sup>.



**Figure 3 (Adapté de 6 et 7)- Représentation schématique de l'évolution des infections par le VHC**

### 4) Cas particulier des co-infections

Un intérêt particulier est porté aux cas de co-infections virales, notamment la co-infection VHC-VIH touchant 4 à 5 millions de personnes dans le monde<sup>1</sup>. Il s'avère qu'une telle co-infection peut augmenter la charge virale du VHC par rapport à une infection classique et provoque, dans un certain nombre de cas, une progression plus rapide du VIH et des maladies hépatiques liées au VHC. Cette co-infection complique aussi considérablement la prise de traitement anti VIH<sup>8</sup>.

Les risques de co-infection par VHC et VHB sont aussi élevés, provoquant des maladies hépatiques plus virulentes que lors d'infection seule par l'un de ces virus.

## 5) Traitement

Pendant près de 15 ans, la bithérapie alliant IFN $\alpha$  et ribavirine (analogue nucléosidique de la guanosine à large spectre antiviral) permettait la guérison en moyenne de 55% des patients atteints par le VHC mais avec d'importantes disparités dans la réponse au traitement.

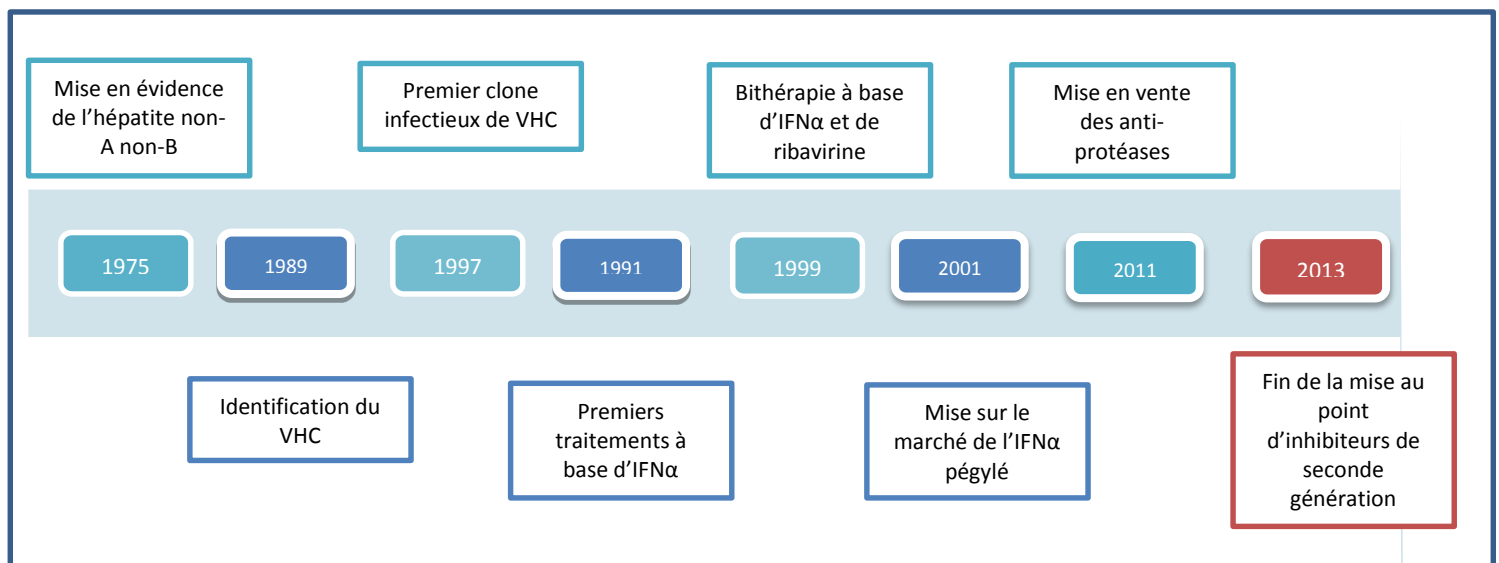
De plus, des effets secondaires importants menaient un nombre trop élevé de patients à arrêter la prise de médicaments.

La mise sur le marché en 2011 d'inhibiteurs de la protéase virale, le Telaprevir et le Boceprevir, a permis un bien meilleur taux de guérison (70-75%) ainsi qu'une prise de traitement plus courte. Cependant, ces protéases étant ajoutées à la bithérapie initiale, leurs effets secondaires étaient cumulés. De plus, le dosage de ribavirine le plus efficace, la procédure de suivi ainsi que la durée du traitement dépendaient d'un certain nombre de paramètres (âge, génotype du VHC, avancée de la maladie...), rendant ainsi l'administration encore plus complexe et la guérison encore plus difficile<sup>9, 10</sup>.

Depuis peu, de nouveaux agents anti-viraux directs (DAA) de seconde génération sont disponibles. Cette avancée est révolutionnaire car ces agents permettent dans un grand nombre de cas la prise seule des DAA, sans prise systématique d'IFN $\alpha$  ou de ribavirine, et s'avère plus efficace que les anti-viraux de première génération. Le traitement est bien mieux toléré et l'estimation du taux de guérison s'élève à plus de 90% sur les essais de phase II et III, tous génotypes confondus<sup>10, 11, 12</sup>.

L'une des molécules les plus efficaces est proposée à environ 50 000 euros pour un traitement complet<sup>13</sup> soulevant ainsi de nouvelles questions plus économiques et éthiques que scientifiques.

Le nombre non négligeable de personnes n'étant pas conscientes de leur infection nécessite, au-delà d'un traitement efficace, le développement d'un vaccin. Malheureusement, sa mise au point reste difficile de par la grande variabilité génétique du virus. C'est pourquoi il est important, entre autre, de continuer à mener des campagnes de prévention anti VHC importantes.



**Figure 4-** Chronologie des événements marquants la mise au point du traitement contre le VHC

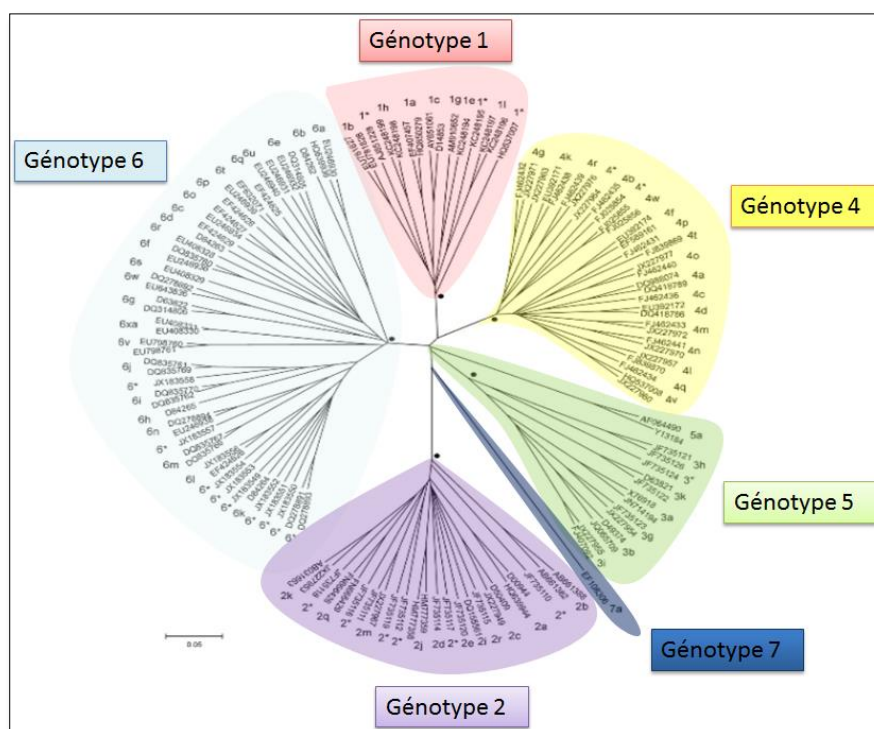
## B. Hétérogénéité génétique du VHC

### 1) Classification

Le virus de l'hépatite C appartient à la famille des *Flaviviridae*, regroupant des virus enveloppés à ARN simple brin de polarité positive; et au genre *Hepacivirus*.

Le VHC est lui-même subdivisé en 7 génotypes différents définis sur des bases phylogénétiques et des analyses de séquences du génome viral entier. L'identité de séquence entre génome complet de deux génotypes se situe entre seulement 50 et 70% <sup>15</sup>.

Les génotypes sont eux-mêmes divisés en sous types (plus d'une centaine) dont 7 principaux : les sous-types 1a, 1b, 2a, 2b, 3a, 4a, 6a <sup>16</sup>. Il existe 70 à 85% d'identité de séquence entre les génomes entiers de deux sous-types du même génotype<sup>15</sup>. Comme la grande divergence des séquences, le nombre conséquent de sous-types répertoriés reflète la grande variabilité génétique de ce virus.



**Figure 5 (adapté de 17)-** Arbre phylogénétique construit à partir de 129 séquences génomiques complètes représentatives du VHC

## 2) Hétérogénéité géographique du virus

Les sous-types du VHC sont épidémiologiquement distincts, ciblant des groupes à risque différents dans des zones géographiques différentes<sup>16</sup>. C'est pourquoi il est important de réaliser une épidémiologie complète du VHC, dans le but d'adapter au mieux les campagnes de prévention ainsi que les traitements à l'hétérogénéité géographique du virus.

La prévalence du VHC est estimée entre 0,1% et 1,5% dans les pays industrialisés et peut atteindre 15% dans des cas similaires à celui de l'Égypte.

Globalement, le génotype 1 est le génotype le plus présent de par le monde et est responsable de 46 % des cas. Les génotypes 2 et 3 sont plus présents dans les pays industrialisés ainsi qu'en Afrique du Sud alors que les génotypes 4 à 6 sont en majorité répertoriés en Afrique et Asie de l'Est<sup>11,18</sup>.

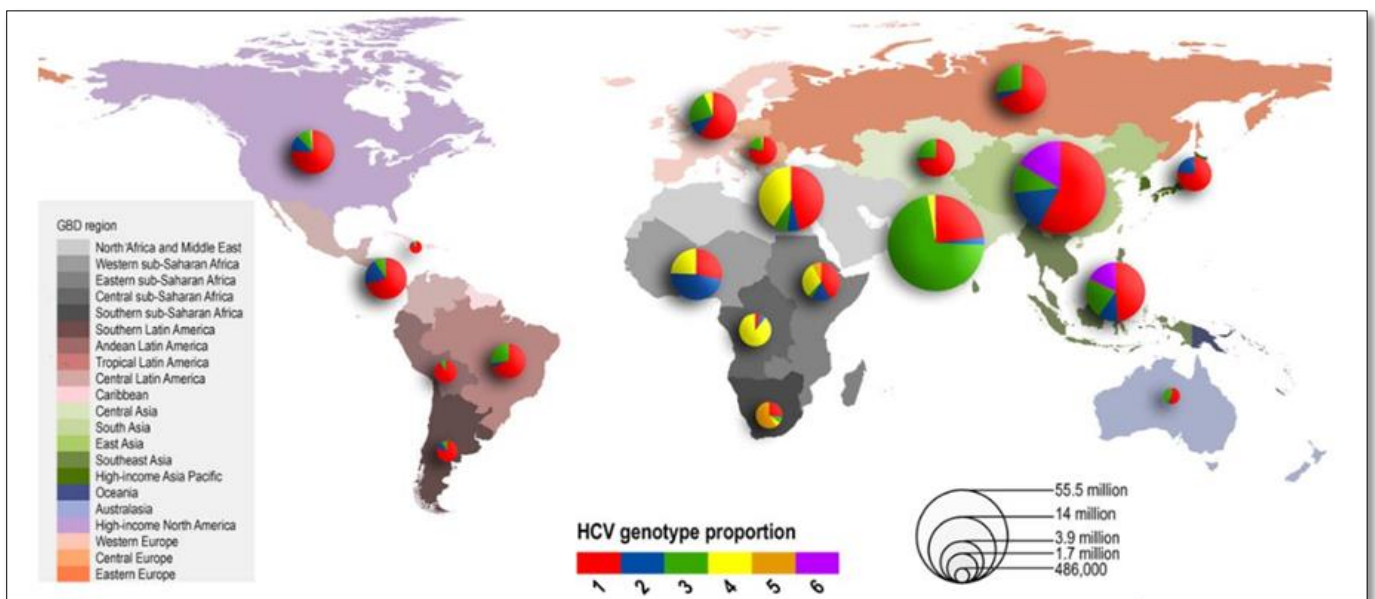


Figure 6- Répartition des génotypes 1 à 6 du VHC dans le monde en 2014<sup>11</sup>

## 3) Grande variabilité génétique du virus

Le VHC présente une grande variabilité génétique au-delà des génotypes et des sous-types. En effet, chaque isolat, soit une souche de VHC isolé à partir d'un seul patient, est constitué d'un pool de virus appelé quasi-espèce. Cette quasi-espèce est constituée de virus dont le génome diffère de 1 à 5%.

Cette grande variabilité permet une adaptation rapide aux changements environnementaux et leur confère ainsi une résistance importante aux agents antiviraux<sup>15</sup>.

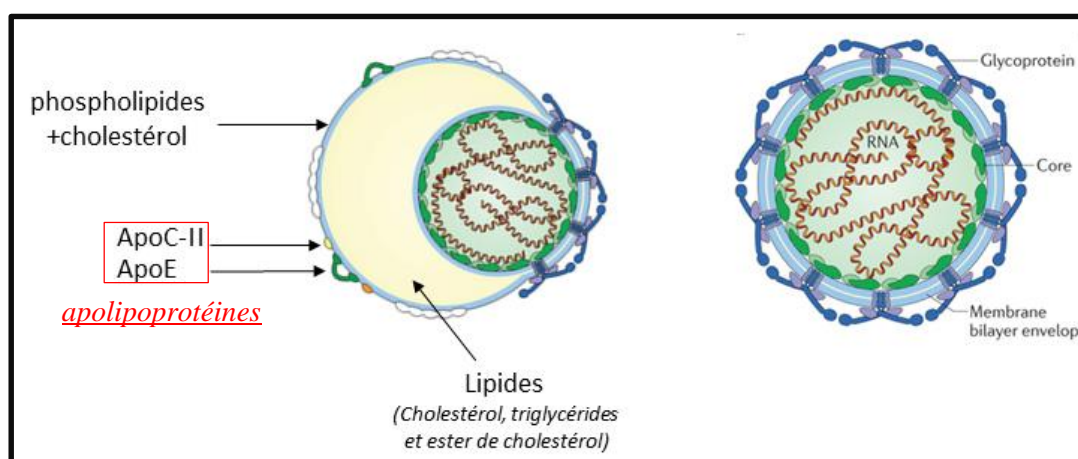
Cette grande variabilité génétique est due au fait que la polymérase NS5B du virus ne possède pas de système de correction. Le taux de mutation est ainsi élevé avec  $10^{-5}/10^{-4}$  nucléotides mutés par cycle de réplication<sup>19</sup>.

## C. Le virus de l'hépatite C

### 1) Structure du virus de l'hépatite C

Le VHC est un petit virus enveloppé d'environ 50nm de diamètre. La particule virale se présente en plusieurs strates<sup>22</sup>:

- La **nucléocapside** formée de la protéine core C entourant une unique copie du génome viral
- L'**enveloppe virale** dans laquelle sont ancrées les glycoprotéines E1 et E2 via leur domaine transmembranaire
- La nucléocapside enveloppée est elle-même entourée d'une couche lipidique formée de phospholipides et de cholestérol dans laquelle sont insérées des apolipoprotéines. Cet ensemble est appelé **particule lipovirale LVP**. Cette enveloppe supplémentaire facilite l'internalisation du virus dans la cellule hôte et protège le virus des anticorps produits par la personne infectée<sup>21</sup>.



**Figure 7 (Extrait de 21)-** Schéma d'une LVP (à gauche) et zoom sur la particule virale enveloppée seule composée de la nucléocapside et des glycoprotéines d'enveloppes E1 et E2 (à droite)

## 2) Organisation du génome viral

Le VHC est un virus à ARN monocaténaire de polarité positive. Il possède une ORF entourée de régions non traduites (NTRs) qui sont essentielles à la bonne réplication du génome<sup>18</sup>.

Le 5'NTR est composé de 341 nucléotides localisés en amont de la région codante et est composé de 4 domaines hautement structurés de type tiges-boucles ou pseudo-nœuds. Cette séquence nucléotidique inclut aussi un site d'entrée interne pour le ribosome (IRES) qui initie la traduction coiffe-dépendante du génome du VHC en recrutant des protéines virales et cellulaires comme le facteur d'initiation eucaryote EIF2<sup>23</sup>.

L'ORF principale de ce génome code pour une polyprotéine d'environ 3000 résidus. Après clivages co- et post-traductionnels, 10 protéines réparties en 2 classes sont obtenues :

- Le **module d'assemblage**

Il comprend la protéine core C ainsi que les glycoprotéines E1 et E2 dites structurales ainsi que les protéines p7 et NS2. Ce module permet l'assemblage et la formation des capsides virales.

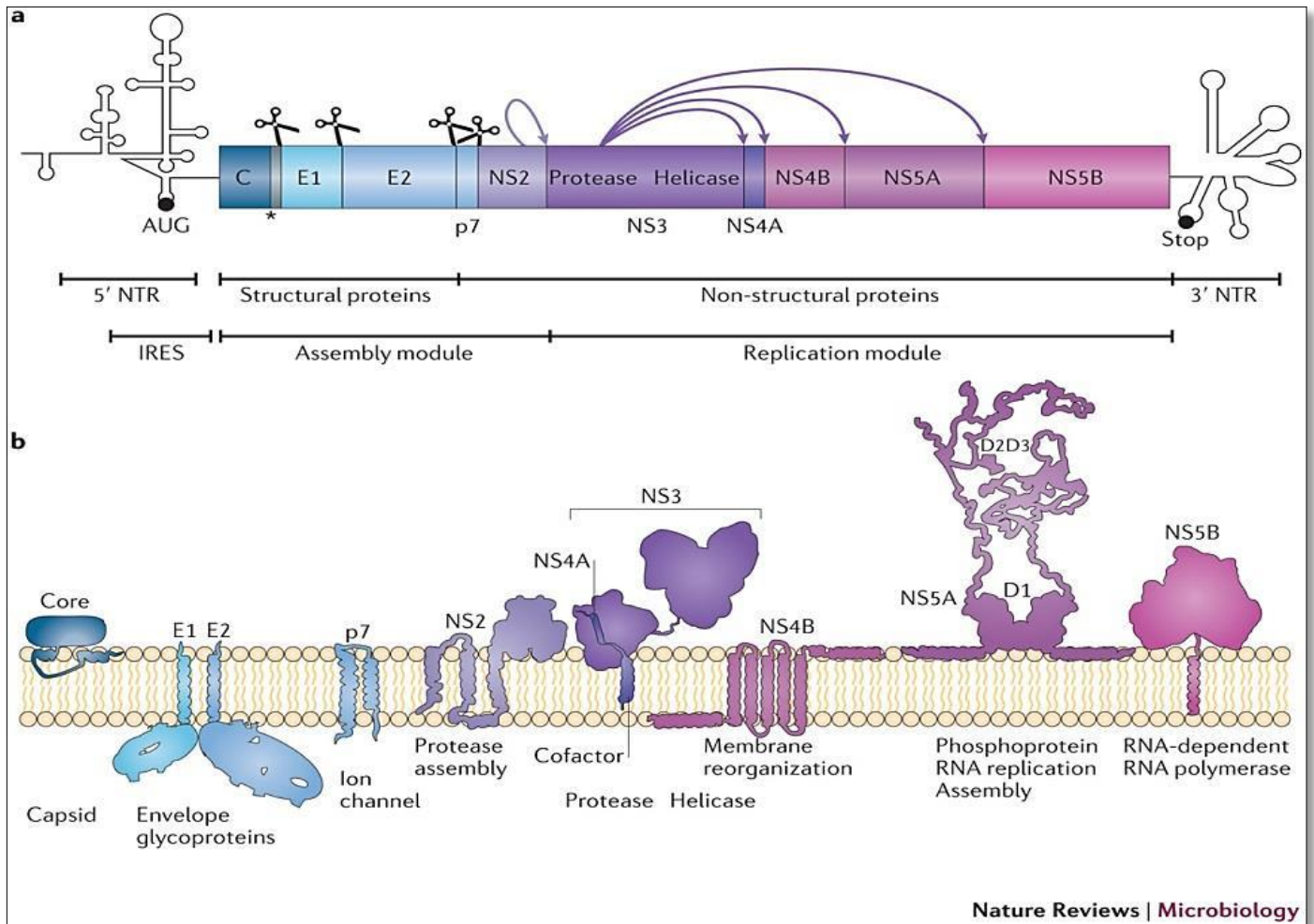
- Le **module de réplication**

Ce module est composé de protéines dites non-structurales : NS3, NS4A, NS4B, NS5A, NS5B. Elles forment le complexe de réplication du génome viral.

L'organisation au sein de la membrane du réticulum endoplasmique de ces protéines ainsi que leur fonction sont repris figure 8.

A noter qu'une petite protéine supplémentaire, la protéine virale F, est traduite à partir de la région codant pour la protéine core C par décalage du cadre de lecture. Sa fonction biologique n'est pas encore définie. Cette protéine ne serait pas essentielle à la réplication du virus<sup>24</sup>.





**Figure 8-** (tiré de la référence 18)

(a) ORF codant pour la polyprotéine du VHC. Les ciseaux représentent les sites de clivage reconnus par les peptidases-signal du réticulum endoplasmique alors que les flèches représentent les sites de clivage des protéases virales soit la protéase à cystéine NS2-3 et le complexe NS3-4A de type protéase à sérine.

(b) Schématisation des différentes protéines du VHC au sein de la membrane du réticulum endoplasmique ainsi que leurs fonctions majeures. A noter que toutes les protéines sont ancrées à la membrane.



### 3) Cycle viral du VHC

- Entrée du virus dans la cellule hôte

Il existe probablement plusieurs voies d'entrée du VHC et seulement l'une d'entre elles a été validée. Elle est inhabituellement complexe et comprend de nombreux corécepteurs. La particule virale va tout d'abord se lier au récepteur scavenger SRB1 ainsi qu'à CD81, une petite protéine de la famille des tétraspanines. Cette interaction va permettre l'interaction avec les protéines claudine 1 et occludine, des protéines de jonctions serrées, permettant l'entrée du virus dans la cellule par endocytose clathrine dépendante.

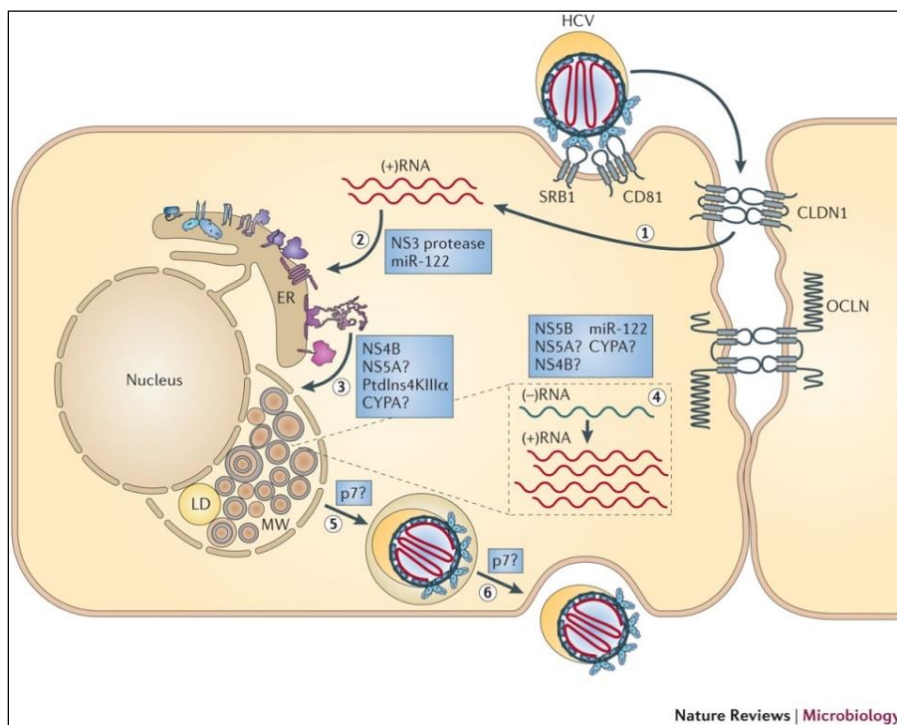
- Traduction et réplication du génome viral

Après internalisation du virus, l'ARN viral est libéré dans le cytoplasme et va être traduit au niveau du réticulum endoplasmique rugueux par la machinerie de traduction cellulaire. La polyprotéine virale va être ainsi synthétisée puis clivée en protéines matures par les protéases cellulaires puis virales. Les protéines virales néo-synthétisées, sans doute en collaboration avec des protéines de la cellule hôte, vont alors induire un fort réarrangement des membranes intracellulaires. Cette déformation est caractérisée par la formation d'un réseau membranaire (membranous web) visible sous la forme de vésicules contigües. Ce réseau membranaire abrite les complexes de réplication virale pour lesquels il est indispensable.

Au sein de ces usines de réplication, la polymérase ARN dépendante NS5B va synthétiser dans un premier temps un ARN intermédiaire à polarité négative qui va servir de brin matrice pour la réplication du génome viral. Les nouveaux brins d'ARN positif sont ensuite synthétisés en excès afin d'être traduits, répliqués ou encore encapsidés puis enveloppés.

- Assemblage et sortie du virus

Les nucléocapsides se formeraient à proximité de la membrane du réticulum endoplasmique et des gouttelettes lipidiques où s'accumulent la protéine core C et l'ARN viral. L'enveloppe virale est acquise par bourgeonnement à la membrane du réticulum endoplasmique via un procédé lié à la synthèse des lipoprotéines. Après maturation du virion, les particules virales vont enfin être libérées par voie sécrétoire.



**Figure 9 (extrait de 18)- Cycle viral du VHC**

- 1- Internalisation du virus par endocytose
- 2- Traduction du génome viral par la machinerie de traduction cellulaire
- 3- Formation du réseau membranaire
- 4- Réplication du génome viral
- 5- Assemblage des particules virales
- 6- Sortie du virus

*CLDN1* = Claudine 1 ; *OCN* = occludine ;  
*LD* = gouttelettes lipidiques ;  
*MW* = réseau membranaire ;  
*ER* = réticulum endoplasmique

## D. Réseau membranaire

### 1) Un mécanisme utilisé chez les virus à ARN positif

Les virus à ARN positif constituent leurs complexes de réplication au sein des membranes d'organelles intracellulaires en provoquant des réarrangements massifs de membrane. Ces environnements favorables au virus sont connus sous le terme d'usines virales ou de réseau membranaire dans le cas du VHC<sup>25</sup>.

Le rôle de ces structures membranaires n'est pas très bien défini mais elles pourraient apporter plusieurs avantages au virus. D'une part, elles permettraient une meilleure coordination des différentes étapes constituant le cycle viral, d'autre part une cinétique optimisée de la réplication du génome et de l'assemblage des capsides virales. En effet, la concentration locale des composants nécessaires à ces réactions est plus élevée dans un environnement restreint que libre dans le cytosol. De plus, la ségrégation dans un compartiment membranaire permettrait de mieux échapper aux défenses antivirales de la cellule hôte<sup>26</sup>.

L'étude de ces compartiments membranaires viro-induits est capitale à la bonne compréhension de ces virus car ils sont essentiels à la réplication de leur génome et donc à leur multiplication.

## 2) Morphologie du réseau membranaire chez le VHC

Le réseau membranaire induit par le VHC est en majorité composé de vésicules à double membrane (DMV) de 150 nm environ. Aux points tardifs de l'infection, l'apparition de tubes à double membrane ainsi que de vésicules multi lamellaires (MMV) est observée. Elle serait une réponse de l'hôte au stress induit par le virus. Toutes ces déformations membranaires se forment à partir du réticulum endoplasmique même et y restent étroitement liées (voir figure 10). Une caractérisation biochimique de ces DMVs purifiées a permis de mettre en évidence une activité réplicase virale permettant de situer les complexes de réplication à l'intérieur ou en surface des vésicules. Il est aussi observé la formation de gouttelettes lipidiques cytoplasmiques (LD) pouvant être considérées comme des sous compartiments du réseau membranaire. La majorité des protéines core C se situe en surface de ces organelles essentielles à l'assemblage des virions.

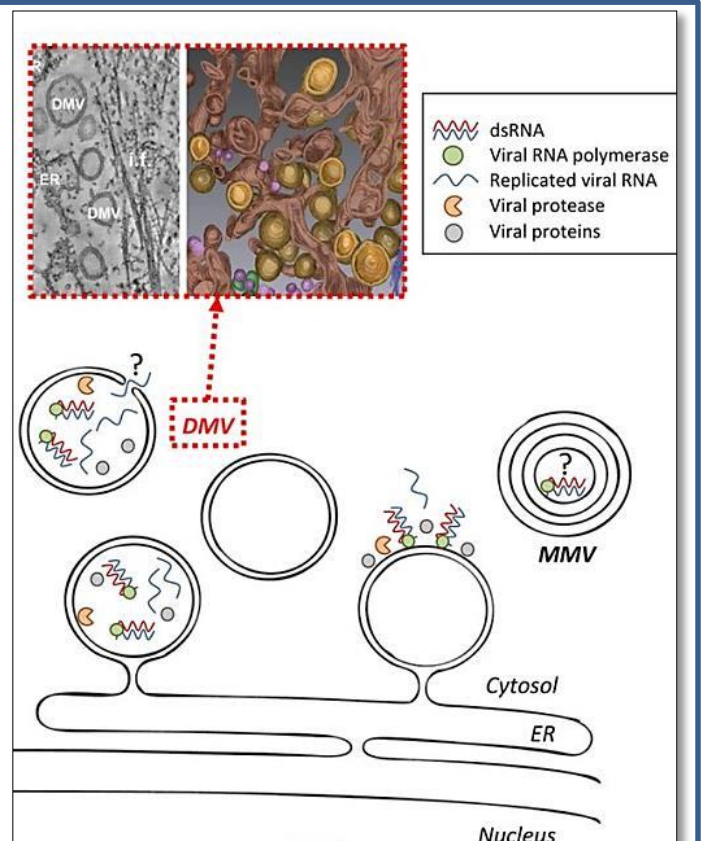
La composition lipidique du réseau membranaire joue un rôle essentiel dans le recrutement des protéines virales et donc dans la formation du complexe de réplication.

Dans le cas spécifique du VHC, il est observé un fort enrichissement de la membrane en phosphatidylinositol-4-phosphate (PtdIns(4)P) et en cholestérol<sup>26</sup>.

**Figure 10 (extrait de 26)** - Représentation schématique de réarrangements membranaires induits par le VHC

En haut à gauche : tomographie obtenue sur une cellule Huh7.5 après 16h d'infection au VHC et reconstruction 3D (obtenue par 27)

DMV = vésicule à double membrane  
MMV = vésicule multi-lamellaire  
ER = réticulum endoplasmique



### 3) Protéines de l'hôte impliquées dans la formation du réseau membranaire

Le VHC détourne un certain nombre de protéines de l'hôte permettant trois changements majeurs dans la cellule, changements indispensables à la formation du réseau membranaire<sup>87</sup>:

#### La restructuration de la membrane du réticulum endoplasmique

La famille des cyclophilines est composée de chaperones catalysant l'isomérisation cis-trans des prolines. Parmi ses 16 membres, un au moins est impliqué dans la réplication du VHC : CypA. Plus particulièrement, Cyp A interagit directement avec les régions désordonnées de NS5A, participant ainsi par un mécanisme inconnu à la déformation de la membrane<sup>93</sup>.

Une seconde protéine de l'hôte, la proline-sérine-thréonine phosphatase-interacting protein 2 (PSTPIP2), est détournée par le virus via son recrutement par NS4B et NS5A<sup>94</sup>. Cette protéine appartient à une famille de récepteurs agissant en tant que senseur et/ou inducteur de la courbure de membrane positive.

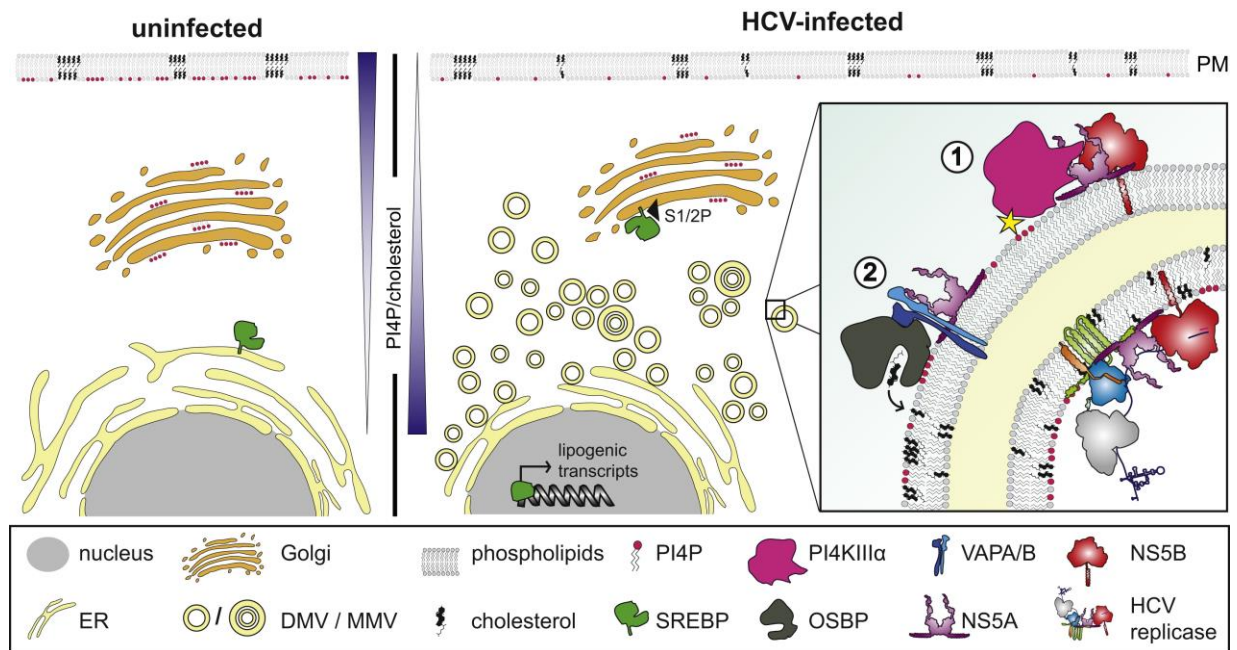
#### Induction de la biosynthèse de lipide *de novo* et de membranes via la régulation de la synthèse de stérols

Lorsque les cellules hôtes sont infectées par le VHC, il est observé le clivage protéolytique de SREBP (sterol regulator element-binding protein) ainsi activé. Cette protéine est un facteur de transcription impliqué dans la synthèse de protéines jouant un rôle dans les voies de signalisation des stérols comme l'acide gras synthase FAS ou encore HMG-CoA reductase (3-hydroxy-3-methylglutaryl coenzyme A), une enzyme à l'origine de la synthèse de précurseur du cholestérol<sup>95</sup>.

#### Changement local de la composition lipidique de la membrane du réticulum endoplasmique

Au niveau du réseau membranaire est observé un fort enrichissement en PtdIns(4)P provoqué par le détournement de la phosphatidylinositol-4-phosphate III $\alpha$  PI4KA sur laquelle je reviendrai plus en détails (Chapitre D.4).

Le PtdIns(4)P est reconnu en particulier par OSBP (oxysterol binding protein)<sup>96</sup> favorisant ainsi le transport du cholestérol vers les DMVs. Cet enrichissement est aussi favorisé par le recrutement de VAPA et VAPB (Vesicle-associated membrane protein-associated protein) par NS5A<sup>97, 96</sup>. En effet, ces deux protéines recrutent OSBP à la membrane du RE (voir figure 11). Un mécanisme similaire à celui du recrutement de l'OSBP par les protéines de recrutement VAPA/B serait engagé dans l'activation de FAPP2 (four-phosphate adaptor protein 2), protéine impliquée dans le transfert de glycosyl-céramides et dépendant du PtdIns(4)P, permettant ainsi l'enrichissement des complexes de réplication viraux en sphingolipides<sup>98</sup>.



**Figure 11** (tiré de 87) - Implication de protéines de l'hôte dans la formation du réseau membranaire du VHC.

**Panneau de gauche :** schéma de cellule naïve ; **Panneau de droite :** schéma de cellule infectée par le VHC. (1) PI4KA interagit avec NS5A et NS5B ce qui permet l'enrichissement de la membrane en PtdIns(4)P. (2) OSBP est recruté à la membrane via le PtdIns(4)P puis va transporter le cholestérol au niveau des membranes à remodeler. Le recrutement d'OSBP se fait aussi via les protéines VAPs.

#### 4) PI4KA : une protéine cellulaire essentielle à la réplication du VHC

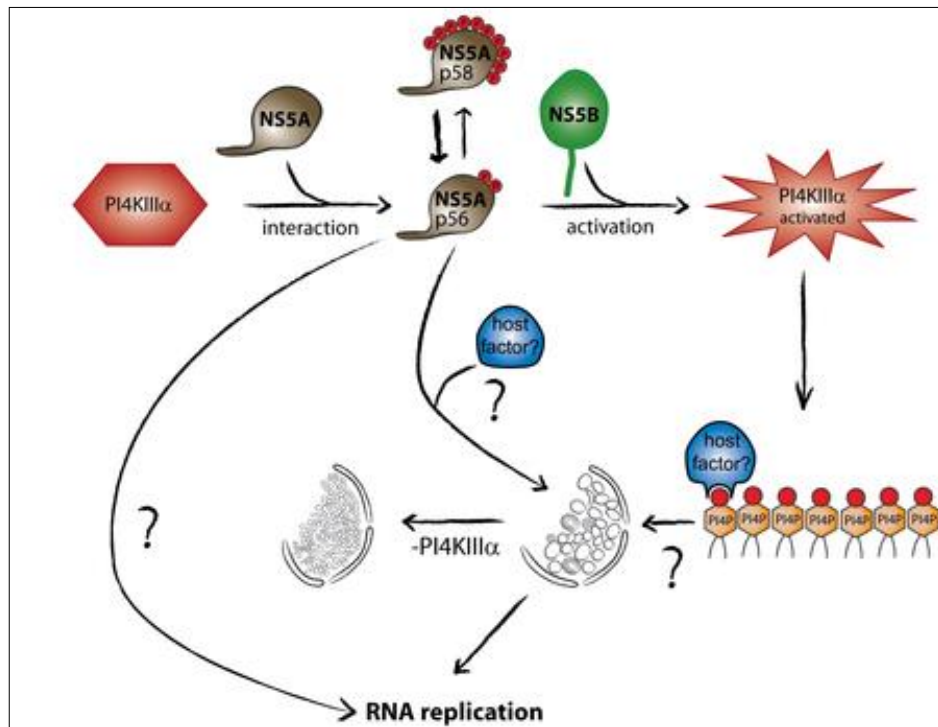
Le fort enrichissement de la membrane du réticulum endoplasmique en PtdIns(4)P lors de l'infection par le VHC est provoqué par le détournement de la kinase de lipide PI4KA par le virus lui-même.

Il a été démontré que cette protéine cellulaire et son activité catalytique sont essentielles à la bonne réplication du génome du VHC. En effet, son silencing au sein de cellules infectées conduit à l'inhibition totale de la réplication du VHC. Il est constaté l'aggrégation des protéines non structurales NS formant le complexe de réplication du génome, ainsi que l'altération de la morphologie du réseau membranaire<sup>30,31</sup>.

Il a été démontré il y a peu que lors de l'infection des cellules par le VHC, PI4KA interagit avec le domaine d1 de NS5A puis est activée par la polymérase NS5B directement ou indirectement. L'accumulation de PtdIns(4)P à la membrane du réticulum endoplasmique est alors rendue possible. Le changement de nature chimique de la membrane, et non pas la forte accumulation de PtdIns(4)P, provoquerait la formation d'un réseau membranaire fonctionnel<sup>87</sup>. En effet, seulement une forte baisse de PtdIns(4)P provoque la formation de DMVs homogènes bien plus petits que ceux observés en temps normal (90nm contre 133 nm en moyenne) provoquant ainsi une abolition de la réplication du génome viral<sup>28</sup>. A noter qu'une déplétion en PtdIns(4)P à la membrane plasmique est aussi observée<sup>28</sup>.



Il s'avère que la protéine NS5A est largement impliquée dans la réplication de l'ARN viral, et plus particulièrement l'un des trois domaines la composant : le domaine d1. NS5A possède deux états de phosphorylation : un état basal p56, favorisant la réplication de l'ARN ; et un état hyperphosphorylé p58, inhibant la réplication de l'ARN et favorisant l'assemblage. D'après des études récentes, PI4KA diminuerait le ratio p58:p56, directement ou indirectement, et aurait donc un double rôle dans la réplication du génome viral<sup>28</sup>.



**Figure 12 (Extrait de 28)**- Schéma hypothétique de la formation du réseau membranaire induit dans les cellules infectées par le VHC.

## II. La phosphatidylinositol-4-kinase III $\alpha$ humaine (PI4KA)

---

### A. Les phosphoinositides : lipides essentiels à la cellule

#### 1) Importances des phosphoinositides

Lors de leur découverte en tant que molécule signal de la membrane plasmique dans les années 70 à 80, les phosphoinositides (PIs) ont su captiver une vaste communauté scientifique. Cette dernière s'est fortement élargie au cours de ces 15 dernières années. En effet, les études menées sur ce sujet ont révélé la place capitale de ces lipides dans un grand nombre de processus.

Les PIs représentent en moyenne 10 à 20% des phospholipides totaux d'une cellule soit une minorité<sup>32</sup>. Pourtant, ils contrôlent à eux seuls en grande partie les processus de vie et de mort cellulaires. Chaque PI possède des fonctions spécifiques réparties en trois catégories générales<sup>32</sup> :

- **Précurseur** dans la synthèse de nouveaux PIs
- **Régulateur** de l'activité et de la localisation de protéines cellulaires, en interagissant directement avec les protéines ciblées<sup>33</sup>. Ainsi, les PIs modulent des canaux ioniques, des pompes, des transporteurs et contrôlent les processus d'exo- et d'endocytose<sup>32,33,45</sup>
- Constituant essentiel de l'**identité des organelles** en apportant une nature de membrane physico-chimique spécifique et donc une signature propre à chaque compartiment<sup>35</sup>. En effet, les PIs sont largement impliqués dans la régulation du trafic vésiculaire, la distribution et le métabolisme des lipides dans la cellule.

De par leurs rôles majeurs dans la cellule, des dérèglements dans le métabolisme des PIs conduisent au développement de cancers, d'obésité, de diabète et de maladies génétiques rares<sup>34</sup>. Des perturbations sont même provoquées par un certain nombre de virus afin d'assurer leur propre réplication<sup>30</sup>. Par conséquent, l'étude des PIs est essentielle tant pour une meilleure compréhension d'un grand nombre de processus cellulaires qu'à des fins médicales et thérapeutiques.

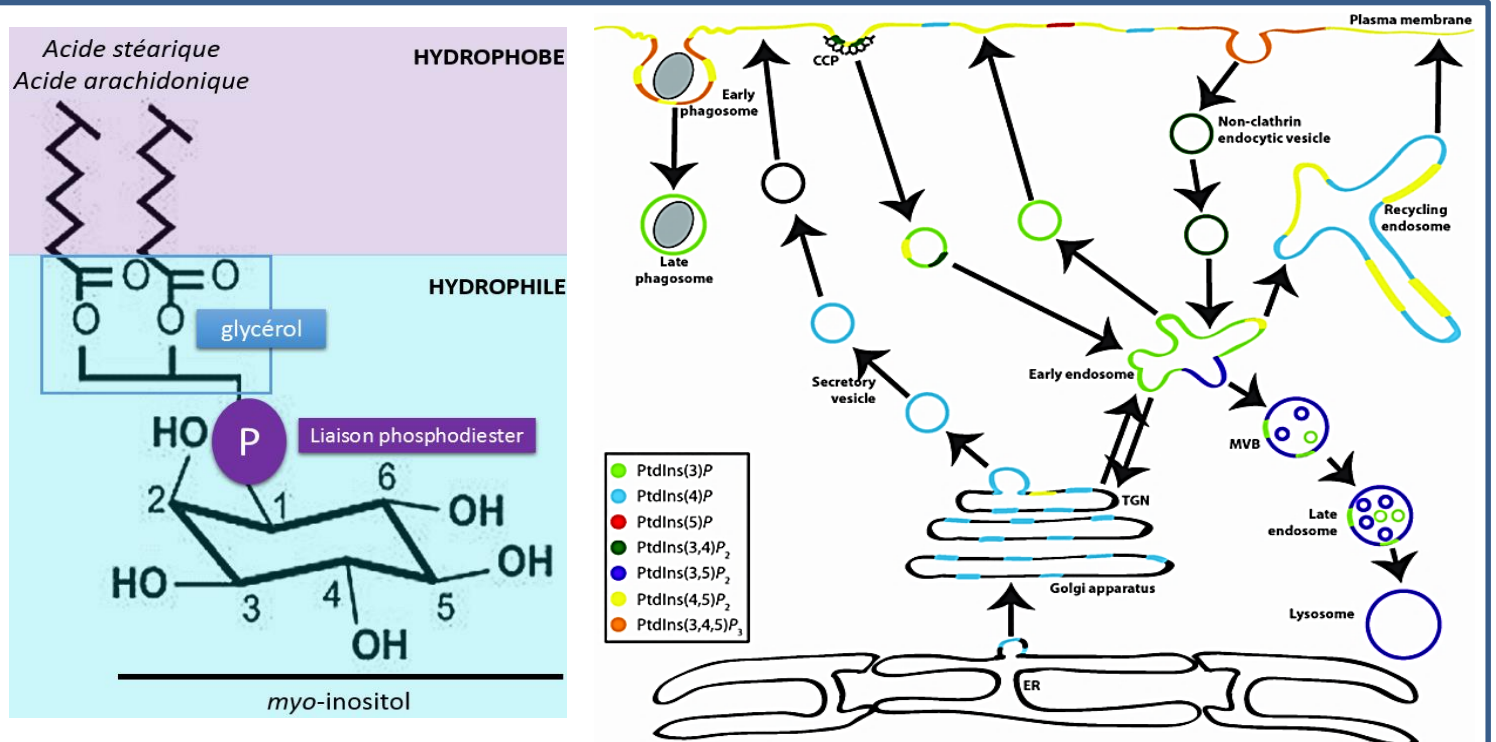
## 2) Nature chimique des phosphoinositides

Les PIs appartiennent à la classe des glycérophospholipides. Ainsi, ils sont formés d'une molécule de glycérol centrale liée par une liaison ester à deux queues lipidiques, communément un acide stéarique et un acide arachidonique<sup>32, 33</sup>. La molécule de glycérol est également associée à un groupement myo-inositol via une liaison phosphodiester en position D1 du cycle (voir figure 13). Certains groupements hydroxyles de l'inositol peuvent être phosphorylés, conférant ainsi au lipide un caractère polaire hautement acide. Les PIs possèdent donc un caractère amphiphile prononcé : ils sont composés d'un groupement hydrophile assez imposant et d'un groupement hydrophobe permettant leur insertion dans la membrane.

Sept PIs naturels, différenciés sur l'état de phosphorylation de l'inositol, sont répertoriés à ce jour :

- Trois PIs mono-phosphorylés : PtdIns(3)P, PtdIns(4)P, PtdIns(5)P
- Trois PIs bi-phosphorylés : PtdIns(3,4)P<sub>2</sub>, PtdIns(3,5)P<sub>2</sub>, PtdIns(4,5)P<sub>2</sub>
- Un seul PI tri-phosphorylé : PtdIns(3,4,5)P<sub>3</sub>

Le nom des PIs est régi par une nomenclature précise, définie sous la forme PtdIns(X)P<sub>y</sub> où X correspond à la position du groupement hydroxyle phosphorylé et y au nombre de phosphorylations. A noter que seulement trois des cinq hydroxyles libres de l'inositol sont phosphorylés naturellement (-3,-4,-5).



**Figure 13-**

(a) Composition chimique du phosphatidylinositol PtdIns (adapté de 33)

(b) Localisation moyenne dans la cellule des différents PIs. La répartition de ces lipides témoigne de leur régulation stricte dans la cellule (tiré de 37)



La synthèse des différents PIs dans une cellule est médiée par une régulation précise et stricte dans l'espace et dans le temps<sup>32,35</sup>. Pour cela, la distribution hétérogène cellulaire des PIs ainsi que l'activité des enzymes à l'origine de leur production sont finement régulées. Ainsi, les phosphatidylinositol kinases (PIKs) et phosphatases, enzymes catalysant la synthèse de PIs respectivement par phosphorylation et déphosphorylation, sont spécifiques à chaque compartiment cellulaire défini par un pool de PIs qui leur est propre.<sup>35</sup>

### 3) Métabolisme des phosphoinositides

#### Métabolisme dit « classique » DAG/IP3

Ce cycle métabolique permet, entre autre, de réguler en partie la composition en PIs de la membrane plasmique et ainsi est impliqué dans un large spectre de voies de signalisation cellulaires<sup>32</sup>.

Dans un premier temps, le PtdIns disponible dans la cellule va permettre la synthèse de PtdIns(4)P et de PtdIns(4,5)P<sub>2</sub>. Le PtdIns(4,5)P<sub>2</sub> joue alors notamment son rôle de messager secondaire dans la mobilisation du Ca<sup>2+</sup>, le renouvellement de PtdIns ainsi que dans la transduction du signal<sup>35,36</sup>. Cette voie métabolique est décrite plus en détail figure 14, signalée en turquoise.

Le PtdIns(4)P ainsi que le PtdIns(4,5)P<sub>2</sub>, tous deux localisés en majorité à la membrane plasmique, constituent les PIs majoritaires de la cellule<sup>32</sup>. Ils sont aussi détectés au Golgi et au niveau de certains endosomes du TGN en quantités moindres<sup>37</sup>.

#### Métabolisme des D3-phosphoinositides

Cette voie de synthèse est régie par des PI3kinases dont elle est dépendante.

Elle mène à la synthèse de moins de 1% des PIs totaux des cellules soit: PtdIns(3)P, PtdIns(3,5)P<sub>2</sub>, PtdIns(3,4)P<sub>2</sub>, et PtdIns(3,4,5)P<sub>2</sub>.

A partir du PtdIns initial est généré du PtdIns3P au niveau des endosomes naissants, lui-même converti en PtdIns(3,5)P<sub>2</sub> au niveau des endosomes destinés à devenir des corps multivesiculaires (MVB)<sup>37</sup>. Enfin, le PtdIns(3,4,5)P<sub>3</sub> va être quant à lui formé à la membrane plasmique, lors de l'activation de PI3kinase par des récepteurs de facteurs de croissance (en particulier l'insuline)<sup>33</sup>. Ce dernier est en grande majorité déphosphorylé en PtdIns(3,4)P<sub>2</sub>.

#### Métabolisme du PtdIns(5)P

La voie métabolique du PtdIns(5)P est peu documentée mais des études récentes suggèrent qu'il joue un rôle de régulateur de la signalisation et du trafic intracellulaire<sup>35</sup>. Cette voie métabolique comprend une nouvelle voie de synthèse de PtdIns(4,5)P<sub>2</sub>.

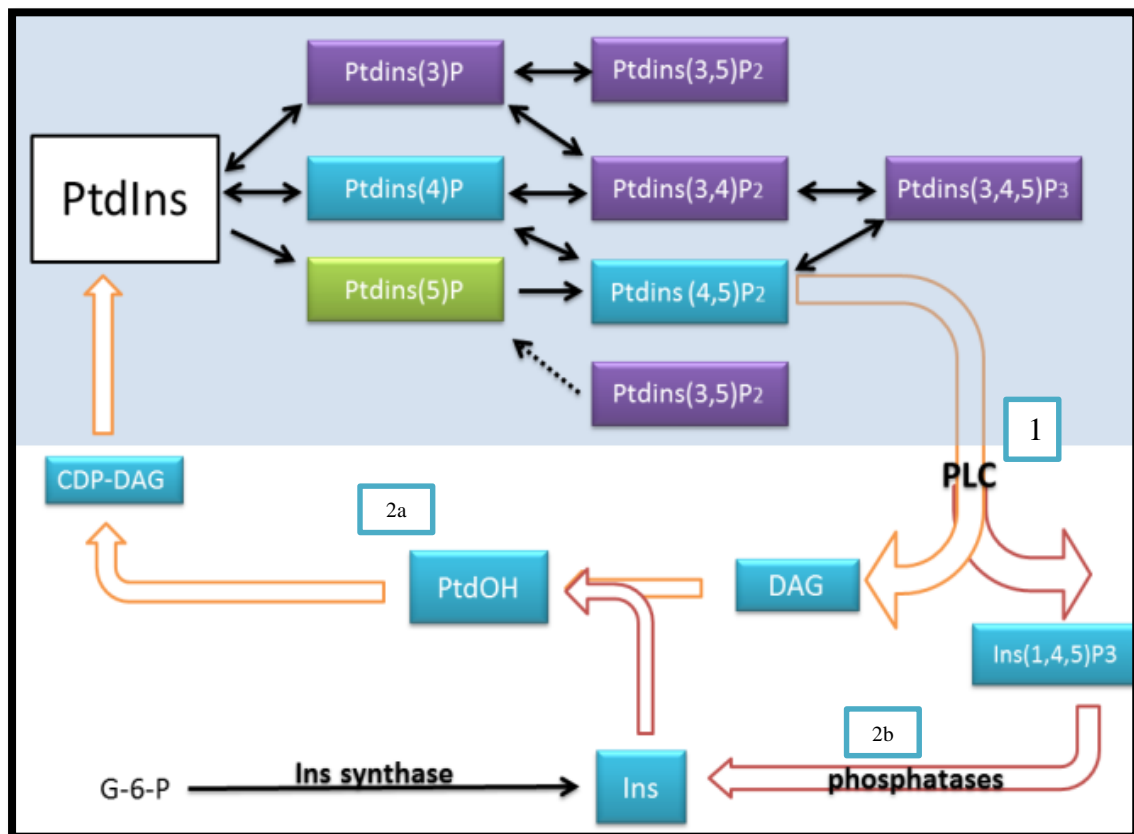


Figure 14 (adapté de 38)- Voies métaboliques des PI

#### Voie DAG/IP3 :

- 1- Production de deux messagers secondaires à partir du PtdIns(4,5)P<sub>2</sub> par une phospholipase C **PLC**: le diacylglycérol **DAG** et l'inositol(1,4,5)-triphosphate **Ins(1,4,5)P<sub>3</sub>**.
- 2a- Conversion du **DAG** en **PtdIns** au niveau de compartiments hautement mobiles dérivés du réticulum endoplasmique puis distribution au sein des autres membranes.
- 2b- Déphosphorylation séquentielle de **Ins(1,4,5)P<sub>3</sub>** en Ins par des phosphatases afin de pouvoir synthétiser du **PtdIns**

#### Métabolisme des D3-phosphoinositides et du PtdIns(5)P

Voies métaboliques régies par des PI-kinases et de PI-phosphatases. Produits de synthèse signalés respectivement en violet et vert

*PtdOH = acide phosphatidique*

*CDP-DAG = Cytidine diphosphate-diacylglycerol*

*G-6-P = Glucose-6-phosphate*

#### 4) Le phosphatidylinositol(4)phosphate

Le PtdIns(4)P est impliqué dans un certain nombre de processus cellulaires de par son rôle de précurseur du PtdIns(4,5)P<sub>2</sub>, mais il est aussi prépondérant dans d'autres événements comme le trafic membranaire ou le métabolisme des sphingolipides.

- **PtdIns(4)P et membrane plasmique**

La majorité du PtdIns(4)P est localisée à la membrane plasmique où il peut être converti en PtdIns(4,5)P<sub>2</sub> et ainsi jouer le rôle de substrat pour la production de PtdIns(3,4,5)P<sub>3</sub>, intervenir dans la mobilisation du Ca<sup>2+</sup> et participer à la transduction du signal<sup>36,39</sup>. Cependant, la majorité du PtdIns(4)P de la membrane plasmique n'est pas essentiel au maintien du pool de PtdIns(4,5)P<sub>2</sub><sup>40</sup>.

PtdIns(4)P ne possède pas seulement des fonctions résultant de la voie classique DAG/IP3 au sein de la membrane plasmique. Ainsi, PtdIns(4)P intervient aussi dans l'ancrage à la bicouche lipidique de protéines comme K-ras, GTPase activatrice de protéines nécessaires à la propagation des facteurs de croissance. Cet ancrage est réalisé par interactions électrostatiques entre la tête acide du lipide et la queue polybasique de la protéine cible<sup>40</sup>.

PtdIns(4)P intervient aussi dans la régulation d'au moins un canal ionique, TrpV1, récepteur activé par une douleur due à la chaleur et la capsaïcine (molécule retrouvée dans les piments)<sup>41</sup>.

- **PtdIns(4)P et membranes internes**

Le PtdIns(4)P est le PI prédominant aux membranes du Golgi et est localisé en quantité moindre dans le réticulum endoplasmique et les endosomes<sup>32,39</sup>.

#### **Complexe Golgien**

Le PtdIns(4)P au sein du Golgi permet essentiellement la régulation de trois classes de protéines effectrices :

- **Les protéines adaptatrices de clathrine**

Certaines vésicules, dites mantelées, possèdent une enveloppe constituée d'un complexe adaptateur qui lie des protéines appelées clathrines aux récepteurs de la vésicule. Cette structure est importante pour la sélection des cargos et par conséquent essentielle au transport de la vésicule. Il existe deux familles majeures d'adaptateurs de clathrine : la famille des adaptateurs AP et GGA<sup>32</sup>.

Les protéines adaptatrices de clathrine adoptent des stratégies similaires. Leur interaction avec le PtdIns(4)P (généré par PI4KIIα) est concomitante à leur liaison avec une petite GTPase Arf1 permettant l'association du complexe à la membrane du Golgi. Ces interactions, combinées à

la reconnaissance du cargo, permettent une fine régulation du recrutement de ces protéines adaptatrices au sein du réseau transgolgien<sup>36</sup>.

- Les protéines de transfert de lipides COF

La famille des protéines de transfert de lipides COF (**Cert**, **OSBP**, **FAPP**) permet le transfert de certaines classes de lipides dépendamment du PtdIns(4)P.

**Cert** assure le transport non vésiculaire de céramides au RE où ils sont convertis en sphingomyéline<sup>42</sup>.

**OSBP** est une sous-famille de protéines assurant pour la plupart le transfert du cholestérol entre le RE et le Golgi. Le mécanisme de régulation de ces protéines via PtdIns(4)P reste obscur. Il existerait aussi un lien entre la relation OSBP-PtdIns(4)P et la synthèse de sphingolipides<sup>32</sup>.

**Fapp1 et 2** sont des protéines impliquées dans le transfert de glycosyl-céramide du TGN à la membrane plasmique<sup>36</sup>.

Toutes ces protéines interagissent avec le PtdIns(4)P via un domaine homologue de la pleckstrine (domaine PH). Tout comme les protéines adaptatrices de clathrines, ce complexe interagit lui-même avec Arf1<sup>32,36</sup>.

- GOLPH3

La protéine de mammifère GOLPH3 lie spécifiquement le PtdIns(4)P et contrôle la morphologie du Golgi. En effet, elle forme des ponts entre la membrane et le cytosquelette d'actine grâce à sa liaison à la myosine MYO18A<sup>32</sup>.

Autres membranes internes

Le PtdIns(4)p est aussi localisé à la membrane du réticulum endoplasmique. Il est généré par deux kinases : PI4KB, synthétisant le PtdIns(4)P intervenant dans la fusion de vésicules COPII avec la membrane du Golgi ; et PI4KA, à l'origine du maintien du pool de PtdIns(4)P à la membrane plasmique<sup>32</sup>. Le PtdIns(4)P des endosomes et lysosomes est quant à lui généré par les PI4KIIs essentiellement, et joue un rôle important dans certaines étapes du trafic cellulaire<sup>32</sup>.

La mise en évidence de PI4KA<sup>73</sup> et B<sup>74</sup> au niveau du noyau suggérerait l'existence d'un cycle de PI nucléaire, mais cela reste sujet à débat.

## B. La phosphatidylinositol-4-kinase de type III $\alpha$ PI4KA

### 1) Les PI4-Kinases

Les PI4Ks phosphorylent en position D4 la tête polaire des PIs. Chez les mammifères, il est répertorié à ce jour 4 isozymes de PI4Ks, classées sur la base de leur activation par des détergents, leur affinité pour l'ATP, l'inhibition potentielle à l'adénosine et par la wortmannin ainsi que leur poids moléculaire. Ces isozymes sont divisées en deux types :

- **les PI4Ks de type II $\alpha$  et II $\beta$** , structurellement distinctes des autres PIKinases
- **les PI4Ks de type III (PI4KA et PI4KB)**, possédant un domaine C-terminal homologue aux PI3Ks. Deux isoformes de PI4KA ont été décrites chez les mammifères. Cependant, l'isoforme 2 serait un variant épissé de l'isoforme 1<sup>43</sup> et serait catalytiquement inactive<sup>44</sup>.

Ces kinases diffèrent entre autre par leur localisation cellulaire. Globalement, les PI4Ks modulent le trafic lipidique inter-organelle, la signalisation des PIs ainsi que le trafic de vésicules intracellulaires grâce à la synthèse de PtdIns(4)P et via des interactions protéines-protéines. Les fonctions majeures de chaque PI4Ks sont répertoriées dans le tableau ci-dessous, tableau adapté de la référence 25.

<i>Nom</i>	<i>Localisation cellulaire</i>	<i>Fonctions</i>	<i>Poids moléculaire apparent (kDa)</i>
<b><i>PI4KII<math>\alpha</math></i></b> <b><i>PI4K55</i></b>	Réseau trans-golgien Endosomes précoces et tardifs	Recrute les adaptateurs clathrine AP1, AP3, et GGAs au TGN Contrôle le devenir des vésicules endocytiques Favorise la dégradation des récepteurs EGF Intervient dans la signalisation Wnt- $\beta$ -caténine	56
<b><i>PI4KII<math>\beta</math></i></b>	Réseau trans-golgien Endosomes précoces et tardifs Translocation à la membrane plasmique	Pas de fonction spécifique connue à ce jour	56
<b><i>PI4KA</i></b> <b><i>PI4KIII<math>\alpha</math></i></b> <b><i>PI4K230</i></b>	Réticulum endoplasmique Golgi (matériel péricentriolaire) Translocation à la membrane plasmique noyau ?	Fournit la membrane plasmique en PtdIns(4)P Régulation des sites de sortie du RE	240
<b><i>PI4KB</i></b> <b><i>PI4KIII<math>\beta</math></i></b> <b><i>PI4K9</i></b>	Golgi Noyau	Régule la sortie de certains cargos du Golgi Intervient dans le transport des céramides médié par CERT au trans-golgi	110

## 2) Rôle cellulaire de PI4KA

### Chez la levure

La majorité des informations sur PI4KA découle d'études sur son orthologue chez la levure Stt4p. Stt4p est localisée à la membrane plasmique où elle maintient le pool de PtdIns(4)P et permet ainsi indirectement la synthèse de PtdIns(4,5)P<sub>2</sub> par la PIP5kinase Mss4p. Cette fonction est essentielle entre autre à l'intégrité de la paroi cellulaire<sup>45</sup>, au bon métabolisme des sphingolipides<sup>46</sup>, et à la bonne organisation du cytosquelette d'actine<sup>47</sup>.

Au-delà de sa fonction essentielle à la membrane plasmique, Stt4p semble jouer un rôle au sein des membranes internes. En effet, Stt4p pourrait pallier aux défauts de transfert d'aminophospholipides entre le Golgi et le RE<sup>48</sup>, renforçant l'idée que la kinase de lipide serait impliquée dans la synthèse de PtdIns(4)P au RE<sup>32</sup>. Stt4p interviendrait aussi dans la bonne formation des lysosomes<sup>47</sup>.

Enfin, de manière plus surprenante, Stt4p est requise pour l'initiation de la traduction<sup>89</sup>.

### Chez les mammifères

Les fonctions cellulaires de PI4KA ne sont pas clairement établies à ce jour. Un des rôles le mieux défini de PI4KA est le maintien du pool de PtdIns(4)P à la membrane plasmique. En effet, une délétion de PI4KA mène à une forte déplétion en PtdIns(4)P de cette dernière et à une compensation de la perte de PtdIns(4,5)P<sub>2</sub> par une hausse de l'expression de PIP5Ks<sup>36</sup>.

PI4KA serait aussi impliquée dans le trafic vésiculaire. En effet, la kinase recruterait le facteur d'échange nucléotidique GBF1 au Golgi via le PtdIns(4)P produit. GBF1 pourrait ainsi interagir avec la membrane et activer Arf1 qui elle-même recruterait COPI et GGAs, menant ainsi à la formation de vésicules mantelées<sup>32</sup>.

PI4KA intervient dans un certain nombre de processus cellulaires importants dont les mécanismes sont peu définis. Ainsi, PI4KA intervient dans l'organisation de l'actine, la régulation du taux de cholestérol à la membrane plasmique, la formation de certaines jonctions RE-PM, l'activation du canal calcique Orai 1<sup>50</sup>...

## 3) Localisation et recrutement de PI4KA

Une étude de 2012<sup>50</sup> a mis en évidence le recrutement temporaire de PI4KA à la membrane plasmique. Ce recrutement implique une séquence d'une cinquantaine de résidus chez l'homme, hautement conservée chez les mammifères, située en N-terminal de la protéine (figure 15a). Ce segment protéique a été longtemps considéré comme non traduit, posant ainsi des problèmes quant à l'étude de PI4KA recombinante en cellule. En effet, il a été longtemps assumé que PI4KA était localisée au niveau de la membrane du RE alors que le PtdIns(4)P qu'elle produisait était détecté au niveau de la membrane plasmique<sup>49</sup>.

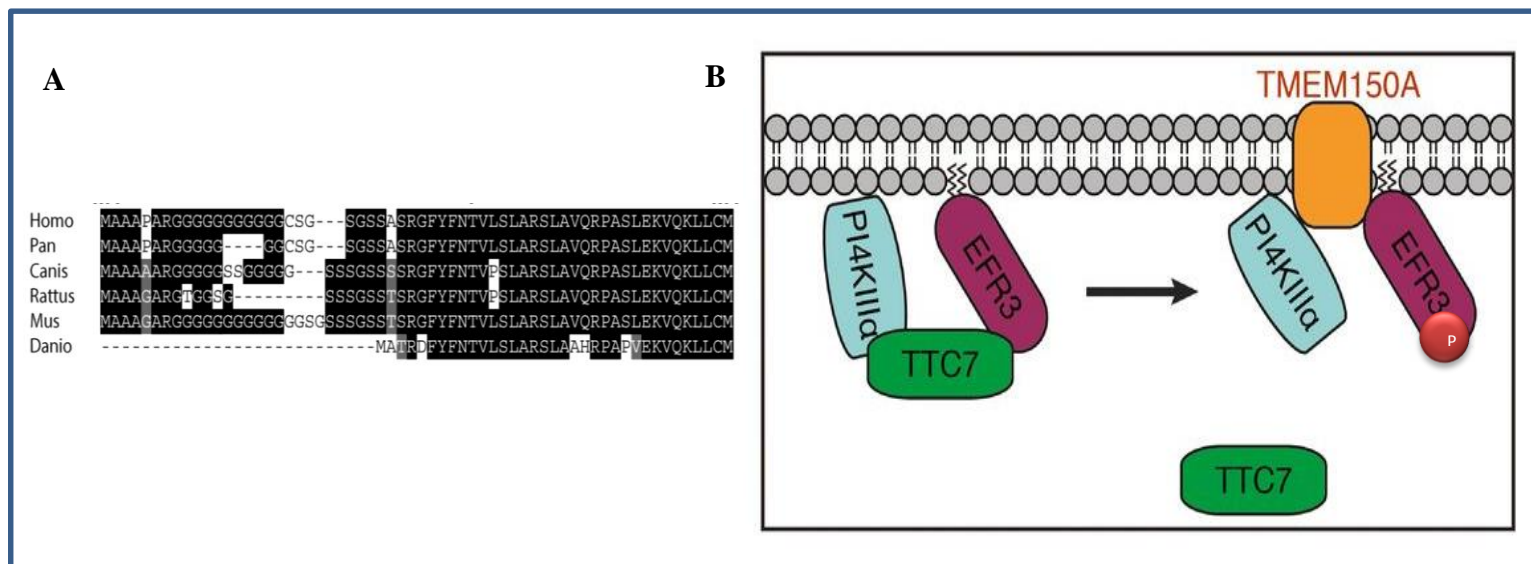
PI4KA est recrutée et transportée du RE à la membrane plasmique par 2 protéines: TTC7 (ou YPP1 chez la levure) et EFR3. Ce complexe, hautement conservé chez les eucaryotes, est responsable environ de la moitié de la production de PtdIns(4)P à la membrane plasmique<sup>50</sup> et est indispensable à la localisation de PI4KA à cette dernière<sup>64</sup>.

**EFR3**, protéine périphérique à la membrane plasmique, permet l'association du complexe à la bicouche lipidique tandis que **TTC7** solidarise le complexe (voir figure 15b). Le recrutement de PI4KA est régulé par la phosphorylation de la partie C-terminale d'EFR3. En effet, EFR3 phosphorylé ne peut pas interagir avec TTC7, empêchant ainsi le recrutement de PI4KA à la membrane plasmique<sup>54</sup>.

Un second complexe, exclusif de TTC7, a été localisé spécifiquement à la membrane plasmique en 2014. Ce complexe comprend PI4KA, EFR3 ainsi que la protéine transmembranaire TMEM150A, située à la membrane plasmique. TMEM150A possède 6 passages transmembranaires ainsi qu'une partie cytosolique interagissant directement avec PI4KA<sup>64</sup>. Le rôle de ce complexe n'est pas défini à ce jour mais il pourrait réguler la production de PtdIns(4)P par la kinase de lipide.

Le taux d'expression des protéines TTC7 et TMEM150A permettent une fine régulation de la localisation intracellulaire de PI4KA. En effet, elles sont dépendantes : plus l'expression de TTC7 dans la cellule est forte, moins il est détecté de TMEM150A et réciproquement. Il existe ainsi un équilibre bien défini entre les deux complexes<sup>64</sup>.

PI4KA a été aussi détectée au sein de l'espace péricentriolaire<sup>51</sup> ainsi qu'au niveau des nucléoles<sup>52</sup>. Ces localisations n'ont été mises en évidence qu'une seule fois et restent donc sujet à débat.



**Figure 15 – Transport de PI4KA à la membrane plasmique**

(a) Alignement des 50 premiers résidus N-terminaux de PI4KA essentiels au transport de la protéine à la membrane plasmique. A noter que cette séquence est fortement enrichie en glycines et sérines (tiré de 50)

(b) Schéma représentant les deux complexes essentiels à la localisation de PI4KA à la membrane plasmique (adapté de 64)



#### 4) Éléments structuraux connus de PI4KA et de ses homologues

PI4KA est une protéine imposante (240kDa-2102 résidus pour la PI4KA humaine) possédant un domaine kinase C-terminal d'environ 400 résidus. La majorité des éléments structuraux connus de PI4KA a été obtenue sur des bases bio-informatiques, plus particulièrement des alignements de séquences<sup>62, 63</sup>. Ces données concernent spécifiquement la région C-terminale de la protéine possédant des homologues de structure connue (résidu 1540 à 2102). En effet, cette région est bien conservée chez les eucaryotes entre PI4KIIIIs et PI3Ks, tout particulièrement le domaine catalytique. Ainsi, il existe 27% d'identité de séquence entre le domaine catalytique de la PI4KA et Vps34, le seul représentant des PI3Ks chez la levure<sup>55</sup>. À noter que VPS34 est le paralogue le plus ancien des 3 classes de PI3Ks chez les mammifères<sup>58</sup>, et représente donc la PI3K la plus proche de l'ancêtre commun entre PI3Ks et PI4KIIIIs. Le domaine catalytique de PI4KA est aussi hautement conservé entre ses orthologues (41% d'identité de séquence avec Stt4p de *Saccharomyces Cerevisiae*) ainsi qu'entre ses paralogues (50% avec PI4KB)<sup>55</sup>.

A contrario, aucune donnée concernant la structure tridimensionnelle et la fonction des 1500 résidus N-terminaux de PI4KA n'est disponible à ce jour. Cette région est assez variable en séquence et aucun homologue (en dehors de PI4KA) n'a encore été détecté. Seul un domaine riche en proline avait été prédit en N-terminal de la protéine<sup>62,63</sup>.

##### i) Domaine Catalytique

Le domaine catalytique de la famille des PI3/4K est composé de deux lobes<sup>59</sup> formant une crevasse dans laquelle est logée la **boucle catalytique** hautement conservée entre les PI3/4Ks. En particulier, le motif 1957–DRH-1959 situé au centre de la boucle est strictement conservé chez PI4KA, de la levure à l'humain<sup>59</sup>.

Les deux lobes sont séparés par une courte chaîne désordonnée comprenant une tyrosine très conservée entre PI3/4Ks (Y1884 pour la PI4KA humaine), tyrosine essentielle à l'activité de PI4KA et de PI3K $\alpha$ <sup>61</sup>. Ce résidu est situé au fond de la poche accueillant l'ATP et fait face à l'adénosine de ce dernier. Aucun mécanisme catalytique n'a été proposé pour PI4KA mais un mécanisme hypothétique de phosphorylation a été proposé pour VPS34 (voir figure 16).

La reconnaissance de la tête inositol du PtdIns ainsi que sa phosphorylation spécifique est possible grâce à la **boucle d'activation**. Cette dernière est de taille variable et est plus ou moins ordonnée suivant les protéines. Elle est hautement conservée seulement entre les PI4KA d'eucaryotes supérieurs<sup>59</sup>.

A ce jour, aucune hypothèse n'a été formulée en ce qui concerne l'interaction de PI4KA avec les membranes. Cependant, chez certaines PI3Ks dont VPS34, il existe une **hélice C-terminale** flexible. La protéine peut ainsi adopter deux conformations : une forme fermée, hélice repliée sur le site catalytique et insérée dans une grande fente, obstruant ainsi le site catalytique ; et une forme ouverte, hélice dirigée vers l'extérieur<sup>58</sup>.

Chez Vps34, cette hélice joue un rôle critique dans l'activité ATPase de la protéine et interagirait potentiellement avec la membrane<sup>58</sup>. Chez PI4KB, les 6 résidus C-terminaux sont quant à eux désordonnés et n'interviennent pas dans la phosphorylation du substrat<sup>57</sup>.



## ii) Domaine à répétitions Armadillo et modules d'interaction

Chez toutes les PI4KIII et PI3Ks, le domaine catalytique est précédé d'un domaine à répétition d'hélices  $\alpha$  de type Armadillo (ARM) possédant des identités de séquence détectables d'une PIK à l'autre. Cependant, ce domaine varie en taille, en séquence ainsi qu'en nombre de répétitions selon la protéine étudiée. Cela peut être expliqué par l'implication du domaine ARM dans les interactions protéine-protéine<sup>56,57</sup>.

En amont du domaine ARM se trouvent des domaines régulateurs impliqués dans les interactions protéine-protéine ou encore protéine-lipides. Par conséquent, ce module contribue au recrutement et à l'activation de PI4KA aux membranes intracellulaires. Les domaines qui le composent sont de nature variable entre PI3/4Ks. Ainsi, le domaine C2 interagissant avec les membranes est souvent observé chez les PI3Ks<sup>32</sup> alors que chez PI4KB, l'ARM est précédé d'une insertion flexible de fonction inconnue et d'un domaine riche en proline intervenant dans la transduction de signal. En ce qui concerne PI4KA, un domaine homologue de pleckstrine (PH) a été prédit en tant que module d'interaction<sup>62,63</sup>. Le domaine PH est connu pour être souvent impliqué dans les interactions protéine-membrane chez les kinases de lipide<sup>91</sup>.

Il s'avère que la construction active tronquée de PI4KA la plus courte exprimée en bactérie comprend les 1230 derniers résidus<sup>60</sup>, soit bien en amont du domaine ARM prédit. Ainsi, un ou des modules de régulation pourraient être positionnés en amont du domaine ARM et seraient essentiels à PI4KA.

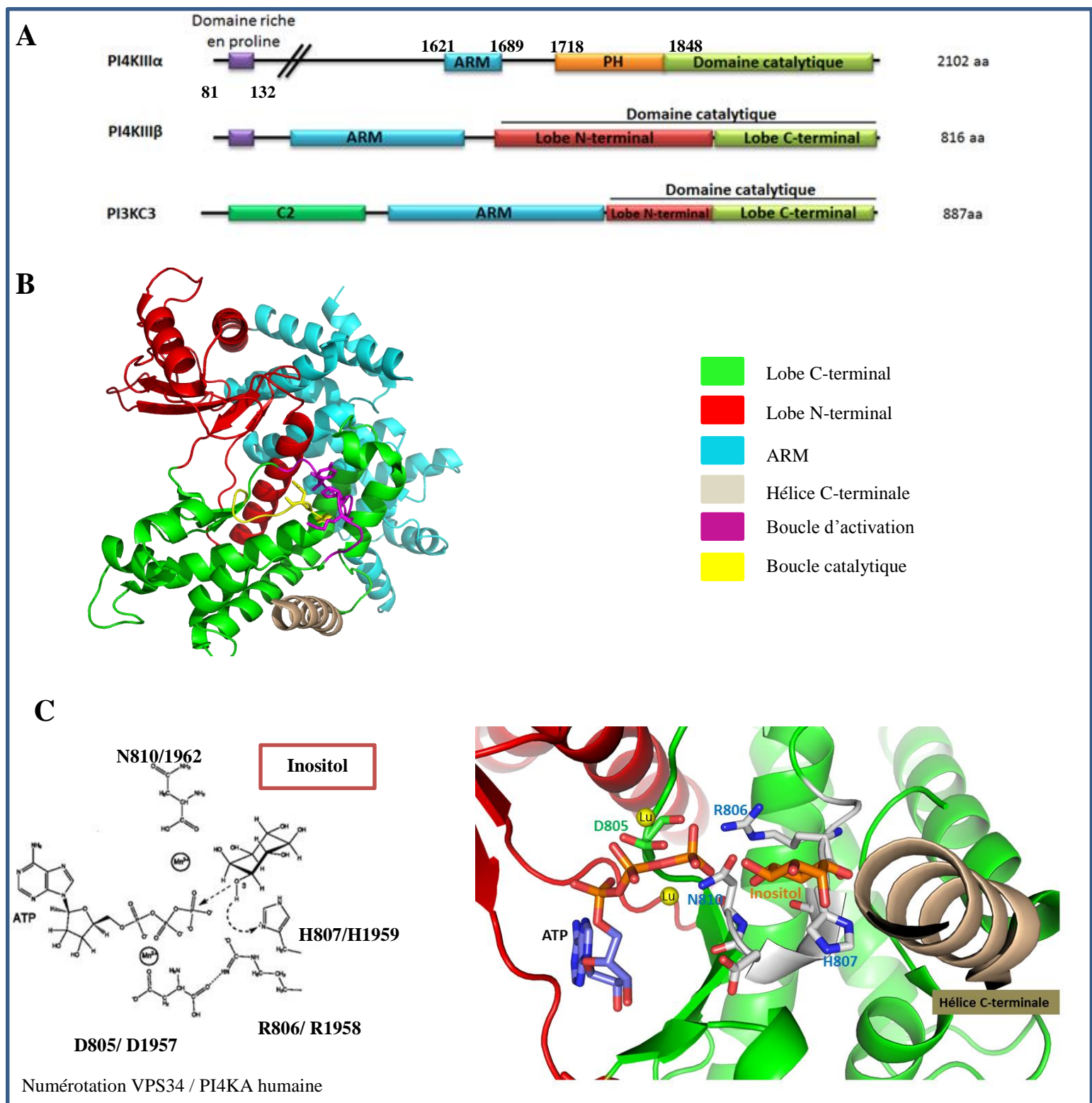
## iii) Cas particulier des Phosphatidylinositol-3 kinase-related kinases (PIKKs)

La famille des PIKKs est constituée de kinases de sérine/thréonine à haut poids moléculaire localisées en périphérie de membranes. Elles ne possèdent pas pour autant d'activité kinase de lipide.

Ces protéines sont constituées d'un domaine C-terminal conservé et homologue aux PI3Ks<sup>79</sup>, et par extension aux PI4KA/B. Cette région de la protéine consiste en un domaine kinase encerclé par un domaine  $\alpha$ -hélical FAT (Focal Adhesion Targeting). Ce domaine FAT, observé en amont du domaine catalytique, peut être composé de différents types de répétitions d'hélices. Ainsi le domaine FAT de la PIKK mTOR est divisé en deux segments  $\alpha$ -hélicaux : le segment TRD, composé de répétitions TPR (tetratricopeptide repeat); et le segment HRD, composé de répétitions HEAT repeats tout comme le domaine ARM des PI3Ks<sup>79</sup>. Ce domaine FAT permettrait la régulation de l'activité kinase des PIKKs<sup>77</sup>.

En C-terminal du domaine kinase est observée une hélice comportant des résidus hydrophobes hautement conservés, critiques pour l'activité catalytique des PIKKs. Cette hélice est proposée comme régulant des interactions protéine-protéine<sup>77</sup> et/ou protéine-membranes<sup>80</sup>.

Les PIKKs se caractérisent par un très long domaine N-terminal flexible (1000 à 3000 résidus) prolongeant le domaine FAT. Cette région de la protéine est entièrement  $\alpha$ -hélicale<sup>77, 79</sup> et forme des super-hélices ou des solénoïdes  $\alpha$ . Peu d'informations sont connues sur ces domaines et sur leur connexion au domaine catalytique, mais ils semblent tenir un rôle majeur dans la régulation de l'activité kinase de PIKKs<sup>79</sup>.



**Figure 16**

a) Organisation en module des protéines PI4KA, PI4KB et PI3KC3 humaines (adapté de 32)

b) Structure cristallographique de VPS34 dans le forme « ouverte » (PDB 2X6F)

c) Hypothèse de mécanisme de phosphorylation de l'hydroxyle D3 du PtdIns par VPS34 (adapté de 58).

Les résidus catalytiques sont strictement conservés entre PI3/4K, c'est pourquoi il a été ajouté à côté de la numérotation de VPS34 celle de la PI4KA humaine.

H807 interviendrait en tant que base catalytique, capturant le proton du 3-OH afin de faciliter l'attaque nucléophile sur le phosphate-γ de l'ATP. D805 et N810 seraient des chélateurs de métaux divalents qui permettraient la neutralisation des charges négatives dans l'état de transition, rôle qui pourrait être aussi endossé par R806. R806 interagirait aussi avec D805 et stabiliserait ainsi la boucle catalytique.

d) Site catalytique de la PI3Kγ humaine (PDB 1E8X) en complexe avec de l'ATP et un inositol.

Les résidus catalytiques sont en représentation sticks et ont été numérotés comme VPS34 pour une meilleure compréhension. La structure a été obtenue en complexe avec de l'ATP et du lutétium, lanthanide susceptible de mimer le  $Mg^{2+}$ . L'inositol a été ajouté à la main via Pymol. A noter que l'hélice C-terminale sur cette structure est repliée sur le site catalytique et est très proche de la boucle catalytique (en blanc)

## 5) Interaction PI4KA-NS5A

### i) Rôle et Structure de NS5A

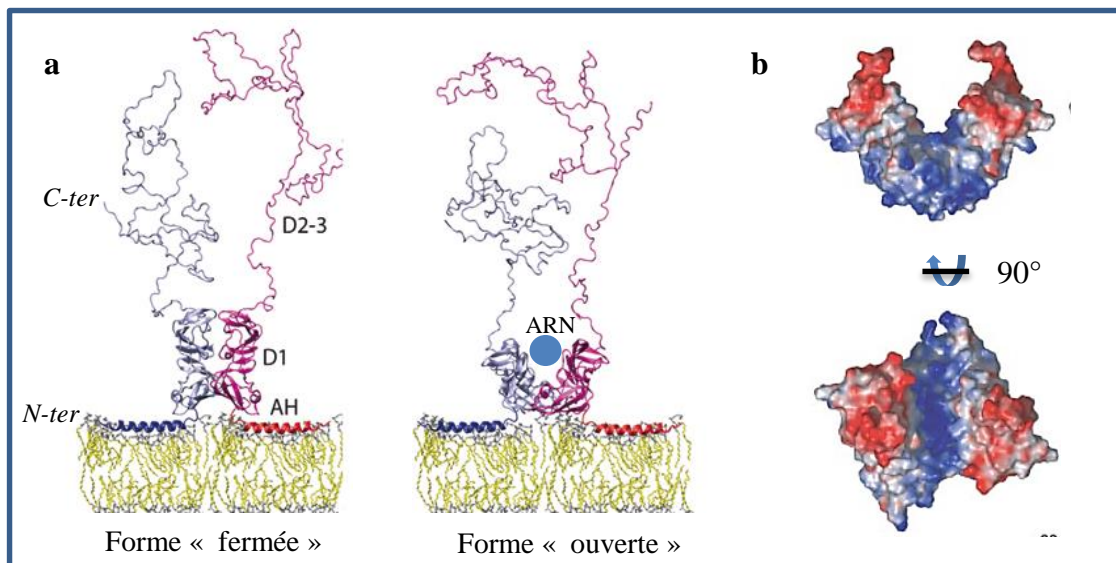
NS5A est une phosphoprotéine essentielle à la réplication du VHC. Elle se compose de 3 domaines (d1-d3) et d'une hélice amphipathique N-terminale permettant l'ancrage de la protéine ainsi que sa localisation aux membranes intracellulaires.

Le **domaine d1**, situé en N-terminal, est essentiellement requis pour la réplication de l'ARN viral. Ce domaine forme des dimères pouvant adopter deux formes alternatives : une forme « ouverte » et une forme « fermée »<sup>66,67</sup>. Ces différents dimères réguleraient l'interaction entre l'ARN viral et la protéine NS5A, et par conséquent la réplication du génome viral (voir figure 17). Il a été suggéré, sur la base de modélisation *in silico*, que le dimère en polymérisant formerait une longue chaîne liant l'ARN viral à la membrane intracellulaire et éventuellement le protégerait de la dégradation<sup>68</sup>. Le domaine d1 possède un motif de fixation au zinc formé de 4 résidus cystéines<sup>65</sup> essentiels à la bonne réplication du virus<sup>65</sup>.

**Les domaines d2 et d3**, produits seuls, sont quant à eux intrinsèquement désordonnés et sous forme monomérique<sup>69,70</sup>. Le domaine d2 est essentiel à la réplication de l'ARN viral<sup>143</sup>, tandis que le domaine d3 joue un rôle principalement dans l'assemblage de particules infectieuses. Il a été répertorié plusieurs phospho-variants de NS5A, dont deux sont prédominants : un état basal de la protéine p56, et un état hyperphosphorylé p58<sup>72</sup>. Le taux de phosphorylation de la protéine virale joue un rôle important dans la régulation de l'assemblage du virion et dans la réplication de l'ARN viral.

NS5A reste une protéine difficile à étudier, que ce soit *in vivo* ou *in vitro*, mais tient une place capitale dans des processus qui semblent complexes. A l'heure actuelle, les mécanismes viraux impliquant la protéine NS5A ne sont pas clairement établis.

Beaucoup de partenaires protéiques cellulaires de NS5A ont été détectés par un large screen en double hybride<sup>71</sup> mais peu d'entre eux ont été validés *in vivo*. L'un de ces partenaires valides est PI4KA<sup>30</sup>.



**Figure 17**

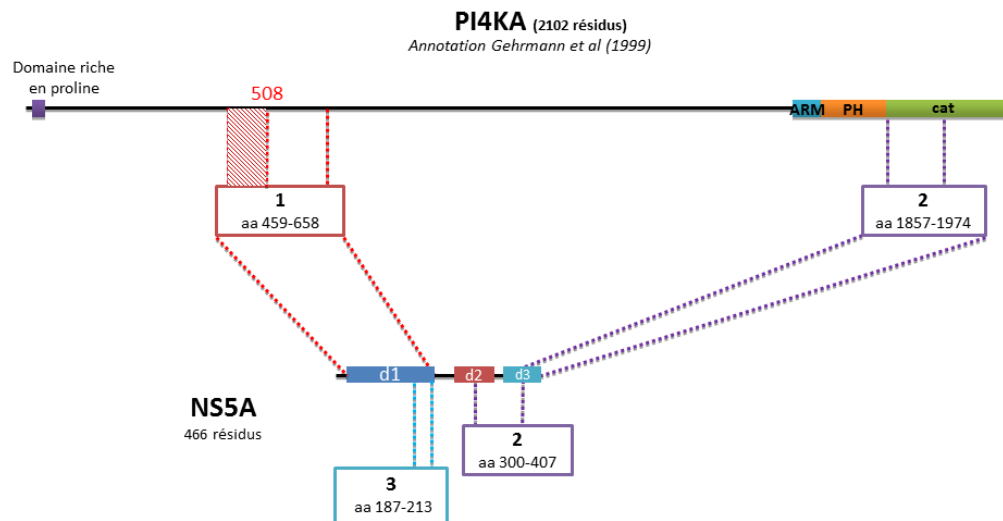
**(a) Modèle des deux conformations adoptées par le dimère de NS5A entière associée à une membrane (adapté de 18).**

Les deux sous-unités du dimère (violette et rose) sont composées des éléments structuraux suivant : l'hélice amphipathique HA (résidus 5-25, PDB 1R7E), le domaine d1 hautement structuré (résidus 33-213, PDB 1ZH1 et 3FQQ), et les domaines intrinsèquement désordonnés d2 (résidus 250-342) et d3 (résidus 356-447).

**(b) Potentiel de surface du dimère « ouvert » formé par le domaine d1 (PDB 1ZH, tiré de 66).** Ce dimère possède un potentiel de surface acide sur la partie haute de la « pince », alors que la partie basse est basique. Cette conformation laisse accès à un large sillon basique susceptible d'interagir avec une molécule d'ARN viral.

## ii) Interaction entre NS5A et PI4KA humaine

Un grand nombre d'études ont montré une interaction directe entre NS5A et la PI4KA humaine. Cette interaction est à l'origine du recrutement de la kinase de lipide au réticulum endoplasmique puis de son activation par la polymérase NS5B. De manière plus spécifique, il est souvent observé une interaction impliquant le domaine d1 de la protéine non structurale. A ce jour, trois études notables proposent une délimitation affinée des sites d'interaction de chacune des protéines :



### 1- Etude Lim et al (78)- publiée en 2011

Ces résultats ont été obtenus par co-immunoprécipitation.

Il a été mis en évidence une interaction entre le domaine d1 de NS5A (résidus 1-213) et différents points situés entre les résidus 459-658 de PI4KA. Une forte interaction est observée plus spécifiquement entre les résidus 458 à 508 de la kinase.

### 2- Etude d'Ahn et al (75)-publiée en 2004

Ces résultats ont été obtenus par double hybride puis confirmé par GST pull down.

Il a été mis en évidence une interaction entre les résidus 300 à 407 de NS5A et les résidus 1857 à 1974 de PI4KA. Ces résultats restent délicats à interpréter car le GST-pull down a été réalisé avec une construction de PI4KA MBP-1795-2102. Cette construction ne comprend donc pas le domaine catalytique en son entier.



### 3- Etude de Reiss et al (28)- publiée 2013

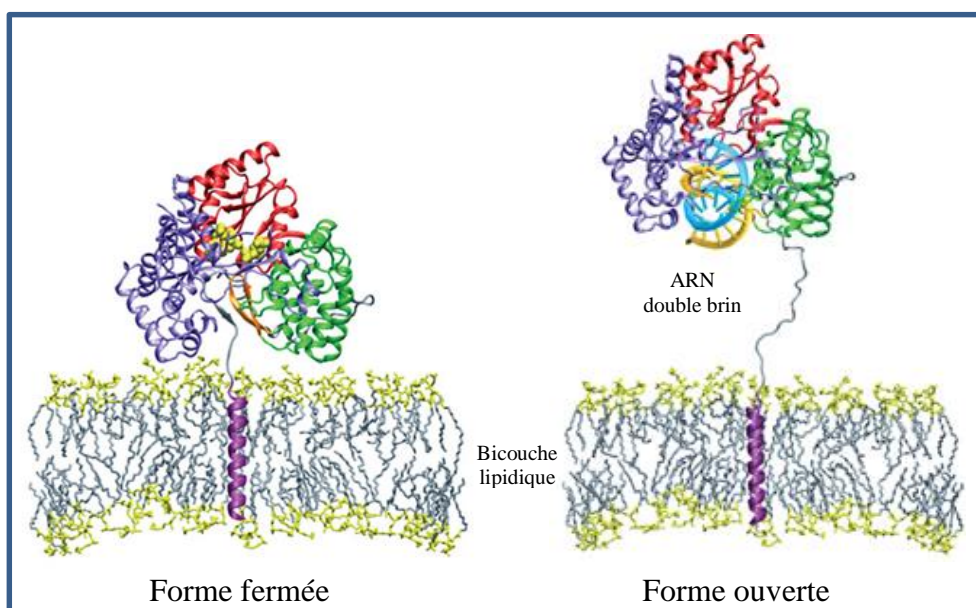
Ces résultats ont été obtenus par co-immunoprécipitation et validés par un grand nombre d'expériences *in vivo*.

Plusieurs sites d'interaction ont été répertoriés dans les derniers 28 résidus du domaine d1 de NS5A (résidus 187-213). Une délétion de cette région induit une déformation des complexes de réplication viraux et une diminution de la production de PtdIns(4)P. Cependant, seulement 25% de l'interaction NS5A-PI4KA est perdue dans le cadre de la polyprotéine NS3-5B, suggérant la présence d'autres sites d'interaction.

Des mutations dans les résidus hautement conservés **203-209** de NS5A abolissent la réplication de l'ARN viral, l'induction de PtdIns(4)P et provoquent une augmentation du ratio p58 :p56. Ce phénotype est aussi observé lors de l'inhibition de l'activité catalytique de PI4KA. A proximité de ces résidus se situe un groupement de sérines essentiel à la synthèse de p58. L'interaction NS5A-PI4KA pourrait ainsi empêcher l'hyperphosphorylation de NS5A et favoriser la synthèse de p56.

### 6) Interaction PI4KA-NS5B

La protéine virale NS5B est une polymérase ARN dépendante permettant la réplication du génome viral. Sa structure s'apparente à une «main droite» composée de trois régions bien distinctes que sont les doigts, la paume et le pouce<sup>83</sup> (voir figure 18). NS5B possède une hélice transmembranaire C-terminale permettant son ancrage dans la membrane. Cette hélice est reliée par un connecteur flexible au reste de la protéine. Ce connecteur joue un rôle important dans la régulation de la réplication de l'ARN viral<sup>84</sup>. En effet, il peut se replier sur le site catalytique de la polymérase permettant ainsi l'accès seulement à de l'ARN simple brin. Il s'agit alors de la forme dite « fermée » à l'origine de l'étape d'initiation de la réplication de l'ARN viral<sup>85</sup>. A contrario, le connecteur est capable de se déplier, libérant ainsi suffisamment d'espace pour une interaction avec de l'ARN double brin : il s'agit alors de la forme dite « ouverte » impliquée dans l'élongation du brin en cours de synthèse<sup>86</sup>.



**Figure 18 (adapté de 18)-** Représentation schématisée de la forme fermée et de la forme ouverte de NS5B en interaction avec un ARN double brin.

Le domaine polymérase de NS5B (PDB 1GX6) sont colorés comme suivant : violet, les doigts ; vert, le pouce ; rouge la paume. Le linker flexible, l'hélice transmembranaire C-terminale de NS5B ainsi que l'ARN double brin de la forme fermée ont été schématisés

Très peu d'éléments sont connus sur l'interaction NS5B-PI4KA. En effet, c'est seulement suite à l'étude de Reiss et al. en 2013<sup>28</sup> que NS5B a été défini comme activateur de la kinase de lipide. Aucun site d'interaction NS5B-PI4KA n'a été défini à ce jour.

### 7) PI4KA, cible potentielle pour un antiviral anti-VHC ?

Qu'en est-il de PI4KA en tant que cible thérapeutique? A ce jour, plusieurs inhibiteurs spécifiques à l'activité catalytique de la kinase de lipide ont été découverts comme par exemple la 4-anilino quinazolin<sup>92</sup>. Cependant, PI4KA tient un rôle physiologique important chez les mammifères. En effet, une souris knock out en PI4KA présente des maladies neurodégénératives ainsi qu'une durée de vie limitée<sup>88</sup>. De plus, PI4KA jouerait un rôle important dans le renouvellement des tissus intestinaux ainsi que dans la division cellulaire<sup>88</sup>. Malgré le détournement de la kinase de lipide par plusieurs pathogènes, comme le VHC et la bactérie gram négative *Francisella Tularensis*<sup>89</sup>, élevant ainsi PI4KA au rang de cible thérapeutique à haut potentiel, il n'en reste pas moins dangereux pour l'organisme de cibler directement cette dernière.

Comme il a été signalé plusieurs fois au cours de cette introduction, PI4KA est impliquée dans de nombreuses interactions protéines-protéines. Cibler des interactions impliquant des protéines virales serait donc une stratégie plus judicieuse et surtout moins dangereuse. Il serait donc plus pertinent de développer des inhibiteurs tels que le daclatasvir, molécule inhibant partiellement l'interaction entre PI4KA et NS5A<sup>90</sup>. En effet, cet inhibiteur bloque NS5A dans la conformation qui serait requise pour le recrutement de PI4KA<sup>81, 82</sup>.

Cette stratégie nous a poussés à étudier la structure tridimensionnelle de la PI4KA humaine afin de mieux comprendre le mécanisme d'interaction PI4KA/NS5A/NS5B. Dans un premier temps, nous avons réalisé des études bio-informatiques afin de délimiter les modules composants la kinase de lipide humaine. A partir de ce découpage, l'équipe de Volker Lohmann à l'université d'Heidelberg a pu définir l'implication de ces modules dans la réplication du VHC ainsi que dans l'activité de la kinase. Leurs résultats surprenants nous ont mené vers la mise au point de l'expression, purification et caractérisation de différents segments de la protéine.

Mes travaux de thèse sont ainsi basés sur une forte collaboration et complémentarité entre différentes équipes de recherche mais aussi sur une interdisciplinarité plus que nécessaire à l'avancement de ces études.



# Partie I

---

Définition des modules et domaines composant  
la PI4KA humaine et fonction potentielle





## I. Définition des domaines C-terminaux de PI4KA par bio-informatique : vers une meilleure compréhension de la réplication du VHC et de la régulation de l'activité catalytique de PI4KA

---

### A. **Article n°1- Mapping of functional domain of the lipid kinase phosphatidylinositol 4-Kinase type III alpha involved in enzymatic activity and hepatitis c virus replication**

#### *1) Ce qu'il faut retenir*

Aucune information récente quant aux domaines composants la PI4KA humaine n'a été publiée dernièrement. C'est pourquoi, dans un premier temps, nous avons décidé de les redéfinir par des méthodes bio-informatiques. Ainsi, nous avons tout d'abord entrepris de délimiter plus finement le domaine catalytique, ARM et PH de la protéine. Une partie des résultats sont exposés dans l'article qui suit.

Le redécoupage de la partie C-terminale de la protéine a été effectué en combinant différentes analyses bio-informatiques:

- Les **alignements de séquences** par PSI-BLAST entre PI4KA et PI3/4Ks

L'homologie entre PI4KA et PI3Ks fut notre plus grand atout. Nous nous sommes donc basés dans un premier temps sur les alignements de séquences pour délimiter les domaines homologues aux PI3Ks que sont le domaine catalytique et ARM. Ces alignements, couplés à une étude des structures à haute résolution des PI3/4Ks, nous ont permis une délimitation fiable des domaines C-terminaux malgré le pourcentage d'identité de séquence un peu faible (30%). La dernière étude de ce type datant de 1999<sup>62,63</sup>, les avancées quant aux alignements de séquences et la parution de nombreuses structures à haute résolution de PI3/4Ks furent des éléments essentiels à la correction du découpage en domaines proposé précédemment.

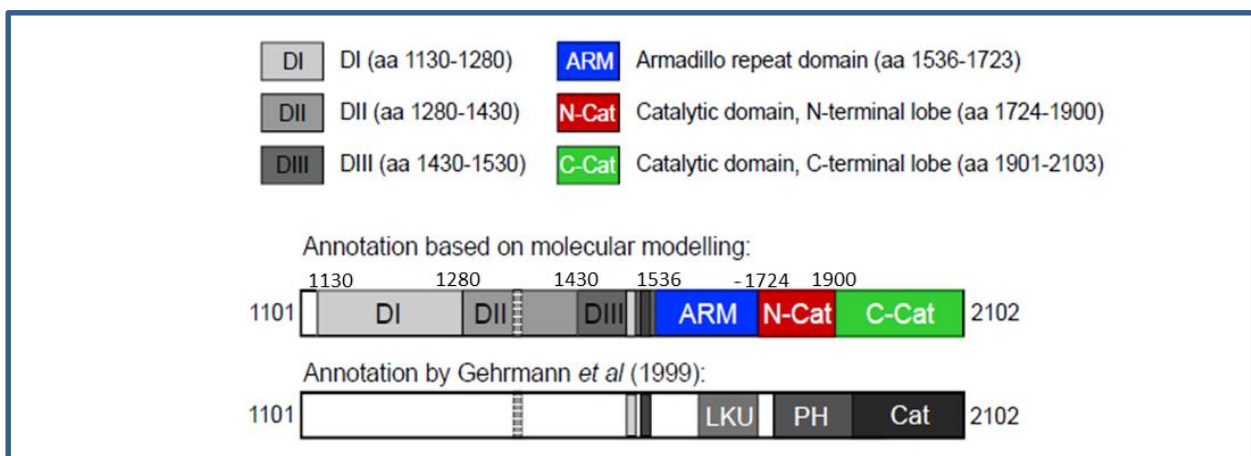
- La **recherche d'homologue lointain** en combinant HHblits<sup>99</sup> et HHPred<sup>100</sup>

La délimitation du domaine ARM a été la plus difficile. En effet, la longueur et la séquence des ARM de PI3Ks étant variables, PSI-BLAST n'est pas assez sensible pour fournir des résultats fiables. Nous avons ainsi utilisé en complément de PSI-Blast HHblits, un programme qui adopte une stratégie similaire mais plus sensible. En effet, HHblits cherche de manière itérative les séquences homologues à la séquence cible dans une banque puis génère des alignements multiples. Cependant, cet alignement multiple est filtré et est basé sur la génération de profils de Markov caché (HMM). Ainsi, HHblits est plus rapide, plus sensible et plus fiable que PSI-BLAST. Afin d'affiner nos études, cet alignement multiple a été analysé par HHPred, logiciel qui recherche les homologues sur la base d'alignement par paire. HHPred étend ses recherches aux banques conventionnelles mais aussi aux banques plus spécifiques comme PFAM<sup>101</sup>.

- Des **prédictions de structures secondaires**

Ces prédictions ont été générées à partir de Quick 2D, serveur regroupant les prédictions de structures secondaires générés par différents logiciels permettant ainsi une redondance des résultats et donc une plus grande pertinence dans l'évaluation et interprétation de ces derniers. Des prédictions de régions désordonnées, importantes pour la délimitation de domaines, ont aussi été réalisées.

Au-delà de la nouvelle délimitation du domaine catalytique et du domaine ARM, nous avons également pu définir trois domaines putatifs que nous avons nommés DI, DII, DIII. Ces domaines ont été délimités seulement sur la base de prédictions de structures secondaires car il n'existe aucun homologue connu à ce jour de cette région de la protéine, hormis PI4KA elle-même. Il s'avère que ces domaines sont prédits entièrement hélicaux. Pour les différencier les uns des autres, nous nous sommes plus particulièrement basés sur la délimitation des hélices prédites ainsi que sur la prédiction de régions désordonnées.



**Figure 19-** Redécoupage en domaine des 1100 derniers résidus de la PI4KA humaine proposé sur la base de nos travaux.

Cette nouvelle délimitation est confrontée au découpage proposé par Gehrmann et al<sup>62</sup> en 1999, et Nagakawa et al<sup>63</sup> en 1996. A l'heure actuelle, ce découpage anciennement proposé est la référence utilisée pour l'étude de PI4KA et le design de mutants par la communauté scientifique.

## - La modélisation par threading

Ces modélisations sont basées sur la relation entre séquence et structure tridimensionnelle ainsi que sur la meilleure conservation de la structure par rapport à la séquence. Nous avons préféré cette stratégie à la modélisation par homologie car PI4KA possède au maximum 27% d'identité de séquence avec les PI3Ks de structure connues, soit la limite pour un modèle fiable généré grâce à la modélisation par homologie.

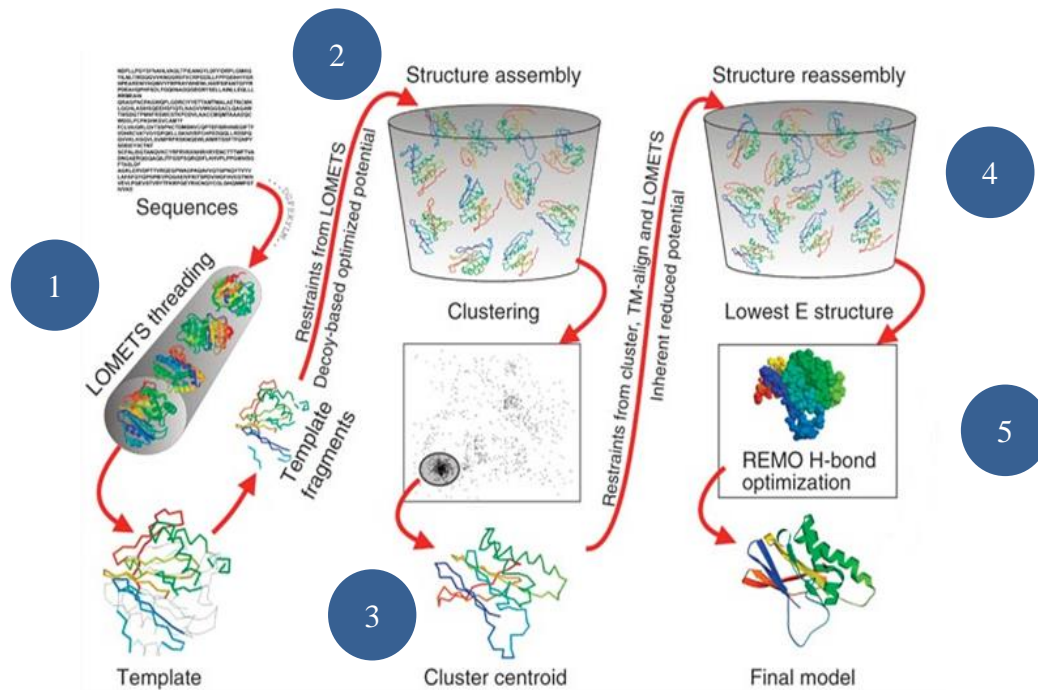
Le modèle présenté dans l'article a été généré via **I-tasser**<sup>102</sup> (Threading, ASsembly, Refinement), programme choisi entre autre pour sa capacité à modéliser *ab initio* les fragments pour lesquels il ne possède pas de structure similaire mais aussi pour sa fiabilité : stratégie itérative, évaluation fiable des modèles, analyse des modèles utilisés, affinement du modèle prédit, mise à jour régulière des méthodes utilisées ... Nous avons choisi de concevoir le modèle publié en imposant un modèle unique, l'homologue VPS34, qui est la PI3Kinase la plus proche de PI4KA et de l'ancêtre commun aux PI3/4Ks. Ce choix a été conforté par I-tasser qui, lors de la modélisation sans contrainte, a évalué VPS34 comme analogue structural le plus proche de PI4KA. De plus, les Confidence Score C-score et le TM score (voir figure 20) obtenus reflètent un modèle plus fiable que ceux générés à partir d'autres modèles uniques (voir tableau ci-dessous).

Deux régions de la partie C-terminale de PI4KA humaine ont été modélisées *ab initio* :

- les résidus 1810 à 1840 du lobe N-terminal du domaine catalytique. Nous avons décidé de retirer ce fragment du modèle publié car il possède une prédiction de structure secondaire peu fiable. En effet, chacun des 30 résidus le composant possède un score inférieur à 4 dans les prédictions réalisées par I-tasser.
- la boucle d'activation (dite aussi "de spécificité"), car cette boucle est connue pour être flexible. Ainsi, cette boucle n'a pas pu être résolue dans la structure cristallographique de VPS34 et des autres structures haute résolution de PI3Ks les plus proches de PI4KA

Pour des raisons pratiques, l'étude approfondie des modèles générés par I-tasser se trouve dans le chapitre suivant (I.A.2), en complément de l'étude des structures cristallographiques des différentes PI3Ks.

A



B

	<i>C-score</i>	<i>TM-score</i>	% identité séquence
<i>Sans modèle imposé</i>	0.16	0.73	NA
<i>VPS34 (PDB 2X6F)</i>	0.80	0.82	23
<i>PI3K<math>\gamma</math> (PDB 1E7U)</i>	0.59	0.79	21
<i>PI4KB (PDB 4D0L)</i>	0.59	0.79	24

**Figure 20- Modélisation par I-Tasser et évaluation des résultats obtenus**

A- Stratégie de modélisation par threading d'I-Tasser (adapté de 103)

1-Recherche de repliements potentiels à partir de la séquence par LOMETS, un serveur d'approche de meta-threading regroupant 9 programmes de threading avec des stratégies différentes. 2- Assemblages des fragments générés par LOMETS et ceux *ab initio* afin d'obtenir des modèles préliminaires entiers par Monte Carlo 3- Sélection des modèles de plus basses énergie 4- Nouvel assemblage des fragments composants les modèles précédemment sélectionnés avec des contraintes supplémentaires afin d'affiner la topologie globale du modèle et d'éviter les clash stériques. 5- Sélection des modèles à plus basse énergie puis génération des modèles atomiques par REMO par affinement des liaisons hydrogène.

B- Evaluation et comparaison de modèles obtenus par I-tasser

**C-score** : estimation de la qualité du modèle prédit. Son calcul est basé sur la pertinence des alignements des structures utilisées pour le threading ainsi que sur la convergence des paramètres de simulation d'assemblage des fragments. Typiquement, le C-score est compris entre -5 et 2. **TM-score** : estimation de la qualité du modèle en évaluant sa similarité avec la structure de la protéine native. En effet, il a été mis en évidence une relation mathématique très étroite entre le C-score et le TM-score expérimental. Le TM donné par I-tasser est par conséquent un TM théorique. S'il est supérieur à 0,5, la topologie du modèle est correcte. Contrairement au RMSD, l'évaluation du TM reflète la topologie globale et non locale.

### Retombées expérimentales :

Ces études bio-informatiques nous ont permis de définir une nouvelle délimitation potentielle plus fiable que celle jusqu'alors utilisée. En effet, au-delà de la délimitation des domaines, certaines prédictions de Gerhmann et al étaient erronées. Par exemple, la topologie du domaine PH prédit en amont du domaine catalytique n'est pas retrouvée dans nos modèles par threading, ni même dans les structures cristallographiques de PI3Ks. En effet, il manque deux brins N-terminaux ainsi que l'hélice C-terminale du repliement type de domaine PH (voir figure 21). Notre nouveau découpage publié est maintenant disponible comme nouveau guide pour le design de constructions plus pertinentes de PI4KA.

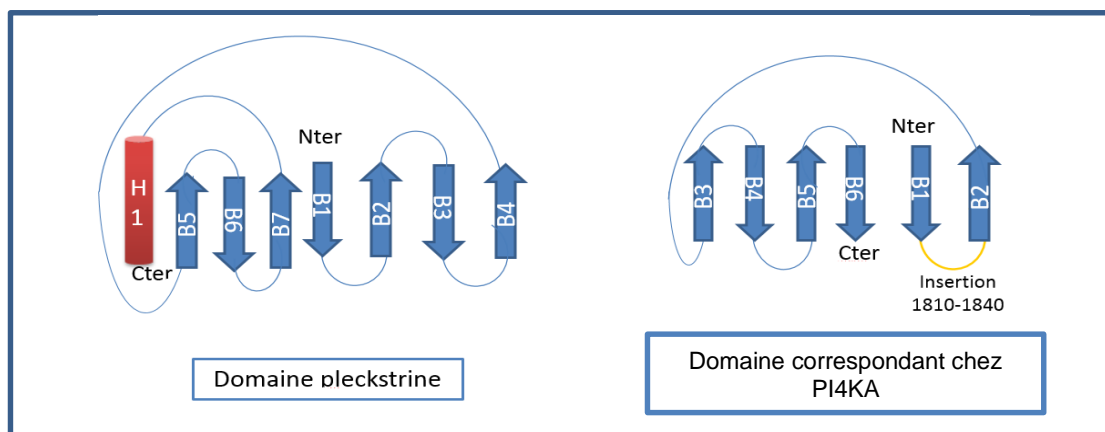


Figure 21- Diagramme de la topologie du domaine pleckstrine et du domaine correspondant chez PI4KA

Cette nouvelle délimitation de la PI4KA humaine a ainsi permis à l'équipe de Volker Lohmann de réaliser des mutants visant à définir le rôle des domaines putatifs DI-DIII dans l'activité catalytique de la kinase de lipide ainsi que leur importance dans la répllication du VHC. De manière plus spécifique, le recrutement de PI4KA à la membrane du réticulum endoplasmique, l'induction de la synthèse de PtdIns(4)P et la bonne répllication du VHC ont été analysés pour chacun des mutants dans le cadre d'expériences *in vivo* dans des cellules d'hépatome infectées (Huh Lunet). L'activité kinase de lipide a elle aussi testée. Les résultats sont résumés dans le tableau qui suit.

Mutants PI4KA	Réplication du VHC	Induction de PtdIns(4)P et activité kinase de lipide	Recrutement au réseau membranaire
<i>sauvage</i>	+	+	+
<i>1102-2102</i>	+	+	+
<i>1253-2102</i>	—	—	—
<i>~Δ DI (1130-1280)</i>	—	—	—
<i>1431-2102</i>	—	—	—
<i>Δ DII (1280-1430)</i>	—	—	—
<i>1531-2102</i>	—	—	—
<i>ADIII (1430-1530)</i>	—	—	—

D'après ces résultats, les domaines putatifs DI, DII et DIII sont indispensables à l'activité catalytique de PI4KA ainsi qu'à la bonne réplication du VHC. Ces domaines pourraient donc être impliqués dans la régulation de la fonction catalytique de la kinase de lipide. De plus, le mutant  $\Delta 1-1101$ , mutant délété de toute la partie de la protéine située en amont de DI, colocalise avec la protéine NS5A dans une cellule infectée au VHC, au contraire du mutant  $\Delta 1-1202$ . Le domaine putatif DI pourrait ainsi être fortement impliqué dans l'interaction avec la protéine virale NS5A. C'est pourquoi nous avons décidé d'étudier DI plus en détail, afin de mieux comprendre le mécanisme par lequel il est essentiel à PI4KA et au VHC. Les résultats obtenus sont détaillés dans la partie II du manuscrit (Partie II. II).

D'autre part, et de manière surprenante, le mutant  $\Delta 1-1101$  est fonctionnel et permet la bonne réplication du virus. Or, nous savons que PI4KA est recrutée à la membrane plasmique grâce à ses cinquante premiers résidus dans les cellules naïves. De plus, dans le cadre de cellules infectées, la protéine virale NS5A piège la kinase de lipide au niveau de la membrane du réticulum endoplasmique. Par conséquent, nous avons posé l'hypothèse que les 1100 premiers résidus de la protéine formeraient un module N-terminal impliqué dans la régulation du transport de la kinase du réticulum endoplasmique vers la membrane plasmique par le biais d'interaction protéine-protéine avec TTC7 et TMEM150A. Un second module, DI-2102, posséderait quant à lui le domaine catalytique et les éléments nécessaires à la bonne régulation de l'activité kinase de lipide de PI4KA. Ce nouveau découpage potentiel en deux modules de la PI4KA humaine à part la suite été vérifié expérimentalement. Les résultats sont exposés dans le chapitre suivant.

Dans l'article qui suit, seulement une petite partie de l'étude bio-informatique de la partie C-terminale de PI4KA est exposée. Les résultats complémentaires seront détaillés dans les chapitres suivants. La majorité des travaux a été réalisée par l'équipe de Volker Lohmann au sein du département de Virologie Moléculaire à l'université d'Heidelberg. Cependant, leurs résultats nous ont permis de diriger la suite de nos recherches. L'article dans son ensemble représente donc un élément clé à la bonne compréhension du déroulement des événements. C'est pourquoi je me suis permise d'intégrer cet article au corps même de mon manuscrit de thèse. Les études et discussion que nous avons nous-même apportées à l'article sont surlignées en jaune.



# Mapping of Functional Domains of the Lipid Kinase Phosphatidylinositol 4-Kinase Type III Alpha Involved in Enzymatic Activity and Hepatitis C Virus Replication

Christian Harak,<sup>a</sup> Danijela Radujkovic,<sup>a</sup> Cytia Tavenau,<sup>b</sup> Simon Reiss,<sup>a\*</sup> Rahel Klein,<sup>a</sup> Stéphane Bressanelli,<sup>b</sup> Volker Lohmann<sup>a</sup>

Department of Infectious Diseases, Molecular Virology, University of Heidelberg, Heidelberg, Germany<sup>a</sup>; Laboratoire de Virologie Moléculaire et Structurale, CNRS UPR3296, Centre de Recherche de Gif, Gif-sur-Yvette, France<sup>b</sup>

## ABSTRACT

The lipid kinase phosphatidylinositol 4-kinase III alpha (PI4KIII $\alpha$ ) is an endoplasmic reticulum (ER)-resident enzyme that synthesizes phosphatidylinositol 4-phosphate (PI4P). PI4KIII $\alpha$  is an essential host factor for hepatitis C virus (HCV) replication. Interaction with HCV nonstructural protein 5A (NS5A) leads to kinase activation and accumulation of PI4P at intracellular membranes. In this study, we investigated the structural requirements of PI4KIII $\alpha$  in HCV replication and enzymatic activity. Therefore, we analyzed PI4KIII $\alpha$  mutants for subcellular localization, reconstitution of HCV replication in PI4KIII $\alpha$  knock-down cell lines, PI4P induction in HCV-positive cells, and lipid kinase activity *in vitro*. All mutants still interacted with NS5A and localized in a manner similar to that of the full-length enzyme, suggesting multiple regions of PI4KIII $\alpha$  are involved in NS5A interaction and subcellular localization. Interestingly, the N-terminal 1,152 amino acids were dispensable for HCV replication, PI4P induction, and enzymatic function, whereas further N-terminal or C-terminal deletions were deleterious, thereby defining the minimal PI4KIII $\alpha$  core enzyme at a size of ca. 108 kDa. Additional deletion of predicted functional motifs within the C-terminal half of PI4KIII $\alpha$  also were detrimental for enzymatic activity and for the ability of PI4KIII $\alpha$  to rescue HCV replication, with the exception of a proposed nuclear localization signal, suggesting that the entire C-terminal half of PI4KIII $\alpha$  is involved in the formation of a minimal enzymatic core. This view was supported by structural modeling of the PI4KIII $\alpha$  C terminus, suggesting a catalytic center formed by an N- and C-terminal lobe and an armadillo-fold motif, which is preceded by three distinct alpha-helical domains probably involved in regulation of enzymatic activity.

## IMPORTANCE

The lipid kinase PI4KIII $\alpha$  is of central importance for cellular phosphatidylinositol metabolism and is a key host cell factor of hepatitis C virus replication. However, little is known so far about the structure of this 240-kDa protein and the functional importance of specific subdomains regarding lipid kinase activity and viral replication. This work focuses on the phenotypic analysis of distinct PI4KIII $\alpha$  mutants in different biochemical and cell-based assays and develops a structural model of the C-terminal enzymatic core. The results shed light on the structural and functional requirements of enzymatic activity and the determinants required for HCV replication.

Worldwide, about 170 million people are chronically infected with hepatitis C virus (HCV). HCV is a positive-strand RNA virus and belongs to the family *Flaviviridae*. The viral genome encompasses about 9.6 kb and codes mainly for a polyprotein of about 3,000 amino acids (aa) that is flanked by nontranslated regions. The polyprotein is cleaved into 10 mature proteins by cellular and viral proteases: core, envelope glycoprotein 1 (E1) and E2, p7, and the nonstructural (NS) proteins NS2, NS3, NS4A, NS4B, NS5A, and NS5B (reviewed in references 1 and 2). The structural proteins core, E1, and E2 are major constituents of the viral particle and, together with p7 and NS2, are involved mainly in the assembly of infectious virions, while NS3 to NS5B are required for replication of the viral genome. NS3 exhibits helicase and NTPase activities in the C-terminal part and an N-terminal protease constitutively bound to its cofactor, NS4A. NS4B is involved in inducing membrane alterations that are required for viral replication (reviewed in reference 1). NS5A is a phosphoprotein critically involved in various host factor interactions and in regulation of viral replication. Two distinct NS5A phospho-isomers have been described, termed p56 and p58, which have been associated with HCV RNA replication and assembly, respectively

(reviewed in reference 3). NS5B is the viral RNA-dependent RNA polymerase.

Viral RNA replication takes place in vesicular membrane alterations designated the membranous web (MW) (4, 5), which consists of double-membrane vesicles (DMVs) and multimembrane vesicles (MMVs) (6–8). Although the precise role of these membrane alterations in RNA replication still is elusive, the appearance of DMVs correlates best with the kinetics of RNA synthesis, and recent reports suggest that these structures indeed harbor the

Received 15 April 2014 Accepted 5 June 2014

Published ahead of print 11 June 2014

Editor: M. S. Diamond

Address correspondence to Volker Lohmann, volker\_lohmann@med.uni-heidelberg.de.

\* Present address: Simon Reiss, Olympus Europa SE & Co. KG, Micro-Imaging Solutions Division, Hamburg, Germany.

C.H. and D.R. contributed equally to this study.

Copyright © 2014, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JVI.01063-14



HCV replication sites (8, 9). The mechanisms behind MW biogenesis still are poorly understood; however, the combined action of all nonstructural proteins seems to be required for MW formation (8). Furthermore, recent studies pointed to a crucial role of the cellular lipid kinase phosphatidylinositol 4-kinase type III $\alpha$  (PI4KIII $\alpha$ ), which has been identified to contribute to the morphology of the MW by interaction with NS5A (7, 10, 11).

PI4KIII $\alpha$  (also called PIK4CA or PI4KA) is an endoplasmic reticulum (ER)-resident enzyme of about 240 kDa in size, and it converts phosphatidylinositol (PI) to phosphatidylinositol 4-phosphate (PI4P). PI4P plays the role of a precursor for synthesis of other phosphatidylinositides, but it is also believed to take part in intracellular trafficking by constituting membrane identity, and several proteins have been identified that specifically bind to PI4P (reviewed in reference 12). PI4KIII $\alpha$  was initially discovered by Gehrmann et al. by screening of a human fetal brain cDNA library (13) and is highly expressed in heart, brain, placenta, skeletal muscle, kidney, spleen, thymus, prostate, ovaries, small intestine, and colon, and it is expressed at low levels in lung, liver, pancreas, testis, and leukocytes (14). Interestingly, in cDNA libraries of various tissues and cell lines, an alternative splice variant of PI4KIII $\alpha$  was discovered that lacked the coding sequence for the N-terminal 1,249 amino acids of the 240-kDa variant. This truncated variant was found to be catalytically inactive *in vitro* and did not support HCV replication, pointing to an important role of residues further upstream in enzymatic function (15) but questioning the general significance of this isoform.

PI4KIII $\alpha$  belongs to the family of mammalian PI4Ks, which comprises two types with two isoforms each (PI4KII $\alpha$ , PI4KII $\beta$ , PI4KIII $\alpha$ , and PI4KIII $\beta$ ) differing in their subcellular localization and being responsible for the synthesis of distinct PI4P pools (reviewed in reference 16). While the yeast orthologue of PI4KIII $\alpha$  is located at the plasma membrane (17), mammalian PI4KIII $\alpha$  mainly resides at the ER (18) but is also located at Golgi membranes (19) and at nucleoli upon exogenous expression (20). The localization of PI4KIII $\alpha$  also differs among cell types, since in neuronal cells and other cell types, like COS-7 or B50 cells, it is prominently found in the nuclei (21). The exact role of PI4KIII $\alpha$  at the ER still is unknown, but it was demonstrated that PI4KIII $\alpha$  is responsible mainly for PI4P synthesis at the ER, the plasma membrane (22, 23), and partly at the Golgi compartment (24).

Type III PI4Ks are structurally related to phosphatidylinositol 3-kinases (PI3K), since the C-terminal catalytic domain and a helical lipid kinase unique domain (LKU) both are highly conserved, and both lipid kinase families share a sensitivity to wortmannin (16, 25). Besides that, little is known about structural determinants of PI4KIII $\alpha$  playing a role in enzymatic function. The N terminus of PI4KIII $\alpha$  has no homology to other known proteins and comprises proline-rich regions and a putative SH3 domain, as determined by structural predictions of different groups (reviewed in reference 14). In the center of the protein, nuclear localization signals (NLS) and leucine zippers (LZ), as well as a helix-loop-helix domain (HLH), are predicted (13, 14); their roles in enzymatic function still are unclear. Also, PI4KIII $\alpha$  has a predicted pleckstrin-homology domain (PH) between the LKU and the C-terminal catalytic domain (13); however, it is not recognized by most algorithms, and it is questionable if this domain is functional (16).

PI4Ks have been found to be important for replication of a number of viruses (reviewed in reference 26). An essential role of

PI4KIII $\alpha$  in particular for HCV RNA replication has been identified by various studies (7, 11, 27–32). While the precise role of PI4KIII $\alpha$  in the HCV replication cycle has not been clarified so far, several studies have shed light on distinct phenotypes modulated by PI4KIII $\alpha$ . Knockdown of PI4KIII $\alpha$  causes a “clustered” distribution of HCV nonstructural proteins in immunofluorescence (IF) (28), which is reflected by an altered ultrastructure of the MW showing DMVs with reduced diameter and lacking MMVs (7). It was observed that PI4KIII $\alpha$  is responsible for the induction of PI4P at intracellular membranes in the presence of HCV in cell culture and *in vivo*, which might be important for MW biogenesis (7, 33). A recent study has shown that increased PI4P levels induced by HCV recruit oxysterol-binding protein (OSBP), thereby shuttling cholesterol to the MW (34), which is a critical component of the DMVs (9). This work mechanistically links enhanced PI4P levels induced by HCV to the changes observed in MW morphology (34). Still, the impact of OSBP knockdown on HCV RNA replication is limited compared to knockdown of PI4KIII $\alpha$ , suggesting that additional mechanisms involving PI4KIII $\alpha$  contribute to HCV RNA synthesis. Recently, it was shown that HCV also causes depletion of PI4P in the plasma membrane, which argues for a general reorganization of PI4P pools in the presence of the virus (24). PI4KIII $\alpha$  directly interacts with NS5A and NS5B (7, 10, 33). Interaction of PI4KIII $\alpha$  and NS5A was shown to be crucial for HCV replication, and the functional interaction site for PI4KIII $\alpha$  on NS5A was mapped to seven amino acids in domain 1 of NS5A (10). Mutations in that region as well as knockdown of PI4KIII $\alpha$  result in increased levels of NS5A p58, whereas overexpression of PI4KIII $\alpha$  favors synthesis of p56 (10). Therefore, the modulation of NS5A phosphorylation by PI4KIII $\alpha$  might contribute to the role of this lipid kinase in HCV replication. The region in PI4KIII $\alpha$  responsible for NS5A binding has been mapped to the N-terminal 400 to 600 amino acids of the kinase by another study (35). However, the functional relevance of this region for HCV replication and PI4KIII $\alpha$  activity has not been elucidated so far.

In this study, we shed light on structural determinants of PI4KIII $\alpha$  that are important for enzymatic activity and HCV replication by providing a detailed characterization of different PI4KIII $\alpha$  mutants. Furthermore, we give novel insights into the organization of C-terminal domains using molecular homology modeling and elucidate their contribution to the role of PI4KIII $\alpha$  in HCV replication.

## MATERIALS AND METHODS

**Cell lines.** The human hepatoma cell line Huh7-Lunet, stably expressing T7 RNA polymerase under zeocin selection, was used for generation of cell lines and transient expression of plasmids encoding HCV proteins that were analyzed in immunofluorescence and immunoprecipitation (IP) assays (36). Huh7-Lunet T7 cells transduced with nontargeting short hairpin RNA (shRNA) or with shRNA targeting endogenous PI4KIII $\alpha$  under puromycin selection were described previously, as were PI4KIII $\alpha$  knockdown cells reconstituted with shRNA-resistant full-length hemagglutinin (HA)-tagged PI4KIII $\alpha$  (10). For reconstitution of PI4KIII $\alpha$  expression, Huh7-Lunet T7 cells with stable knockdown of endogenous PI4KIII $\alpha$  were generated accordingly by transduction of PI4KIII $\alpha$  knockdown cells with lentiviral constructs encoding shRNA-resistant HA-PI4KIII $\alpha$  mutants. PI4KIII $\alpha$  variants used for purification purposes were expressed in Huh7-Lunet cells harboring T7 RNA polymerase under blasticidin selection and transduced with shRNA targeting endogenous PI4KIII $\alpha$ .



**Primers and plasmid constructs.** The lentiviral vector pWPI-BLR (36) was used for cloning of plasmids encoding different HA-tagged PI4KIII $\alpha$  mutants for the generation of stable cell lines under blasticidin selection. The gene encoding the ca. 240-kDa full-length PI4KIII $\alpha$  isoform 1 originally was obtained from Kazusa DNA Research Institute, Chiba, Japan (product no. FXC00322, corresponding to GenBank accession number AB384703; numberings refer to those of this GenBank entry). Notably, this PI4KIII $\alpha$  construct harbors an additional valine residue at the C terminus (position 2103; GenBank entry AB384703) that was not present in other common GenBank entries of PI4KIII $\alpha$  (accession no. NP\_477352.3) but did not affect any function of the kinase (data not shown). Importantly, the PI4KIII $\alpha$  variant encoded by AB384703 is identical to the variant used by Nakatsu et al. (22), containing an N-terminal extension of 58 amino acids compared to sequences of common GenBank entries, comprising 2,103 amino acids in total. These N-terminal 58 amino acids were implicated in dynamic shuttling of PI4KIII $\alpha$  to the plasma membrane (22), and this function should be retained in the PI4KIII $\alpha$  variant used in our study. Notably, amino acid 1 in P42356.3 (13) refers to amino acid 59 in the PI4KIII $\alpha$  variant used throughout this study (AB384703). Construction of pTM and pWPI plasmids, encoding full-length shRNA-resistant HA-PI4KIII $\alpha$ , has been described before (10). PI4KIII $\alpha$  truncation mutants were cloned into pTM and pWPI constructs by PCR on pTM HA-PI4KIII $\alpha$  templates (primer sequences are available upon request). PCR fragments of N-terminal deletions were cloned into pTM plasmids using restriction enzymes NcoI and either KasI ( $\Delta$ 1-1399) or EcoRI ( $\Delta$ 1-1794). The sequence encoding  $\Delta$ 1-1101 was transferred into pTM plasmids from the corresponding pWPI plasmid using restriction enzymes AscI, SpeI, and BstEII. C-terminal deletions were cloned using EcoRI and KasI ( $\Delta$ 1796-2103), SfiI ( $\Delta$ 1447-2103), RsrII ( $\Delta$ 1104-2103), or AccI ( $\Delta$ 802-2103). Construction of the internal deletion  $\Delta$ 802-1399 or  $\Delta$ 1948-1991 was performed by overlap PCR using the following primer combinations: S\_del800\_1400, A\_del800-1400, S\_delC800AccI, and A\_delN1450Kas for pTM  $\Delta$ 802-1399 or S\_DelActive, A\_delActive, S\_DelActiveKas, and A\_EcoStopPI4 for pTM  $\Delta$ 1948-1991. pWPI plasmids were either created by PCR or constructed by transferring PI4KIII $\alpha$  deletion sequences from pTM to pWPI plasmids using the following restriction enzymes: BsiWI, MluI, and XbaI for  $\Delta$ 1796-2103 or BstXI and SpeI for  $\Delta$ 802-2103. pWPI plasmids encoding mutants  $\Delta$ 1-1152,  $\Delta$ 1-1202, and  $\Delta$ 1-1252 were created by overlap PCR, while primers S\_del1-1152\_N and A\_del1-1152\_C were used as flanking primers for PCRs of all three constructs. pWPI plasmids encoding mutants  $\Delta$ DII,  $\Delta$ DIII,  $\Delta$ HLLH,  $\Delta$ NLS, and  $\Delta$ LZ were created by overlap PCR using ScPPT and ALKUSeq as flanking primers. pWPI constructs encoding mutants  $\Delta$ LKU and  $\Delta$ PH were created accordingly by using SPI4K4730seq and A\_ECMV as flanking primers. pTM plasmids encoding mutant  $\Delta$ 1-1152,  $\Delta$ 1-1202,  $\Delta$ 1-1252,  $\Delta$ DII,  $\Delta$ DIII,  $\Delta$ HLLH,  $\Delta$ LZ, or  $\Delta$ NLS were cloned using an NcoI/BstEII fragment from the respective pWPI plasmid, which was exchanged with the NcoI/BstEII fragment yielded from pTM HA-PI4KIII $\alpha$   $\Delta$ 1-1101, while SpeI/BstEII sites were used for cloning of pTM HA-PI4KIII $\alpha$   $\Delta$ LKU or  $\Delta$ PH. Oligonucleotides were purchased from Sigma-Aldrich, and all PCR-derived sequences were confirmed by sequencing (GATC, Constance, Germany). For *in vitro* transcription of HCV subgenomic RNA, the construct pFKI<sub>389</sub>-Lucubi-NS3-3'/JFH1wt $\delta$ g was used as described before (37).

**Transient HCV replication.** Transient HCV RNA replication assays were performed as described previously (38). In brief, five  $\mu$ g replicon-encoding plasmid DNA harboring hepatitis delta virus ribozymes was used for *in vitro* transcription. Purified RNA was transfected in  $4 \times 10^6$  Huh7-Lunet cells by electroporation. Cells were resuspended in 12 ml Dulbecco's modified Eagle medium (DMEM), 2-ml aliquots were seeded into each well of a 6-well plate, and replication was determined by measuring luciferase activity at 4 h, 24 h, 48 h, and 72 h postelectroporation. Values obtained 4 h after transfection were used to normalize for transfection efficiency.

**Immunofluorescence analysis and PI4P quantitation.** For overexpression of HCV or HA-tagged proteins, Huh7-Lunet T7 cells were transfected with LT1 transfection agent (Mirus Bio LLC, Madison, WI, USA) according to the manufacturer's instructions and were fixed 24 h posttransfection. The immunofluorescence protocol was performed as described elsewhere (7). In brief, cells were fixed in 4% paraformaldehyde (PFA) for 20 min and permeabilized with 50  $\mu$ g/ml digitonin for 5 min for imaging of intracellular PI4P or with 0.5% Triton X-100 for 15 min for colocalization analysis of HA-tagged proteins with HCV NS5A. Primary antibodies were incubated in 3% bovine serum albumin (BSA) for 1 h at room temperature (RT). NS5A was detected by using either NS5A-specific monoclonal mouse antibody (9E10; a generous gift from Charles M. Rice) with a final concentration of 3  $\mu$ g/ml or a polyclonal NS5A rabbit antiserum (no. 4952 [36]) at a dilution of 1:500. PI4P was stained using monoclonal mouse IgM anti-PI4P antibody (Z-P004; Echelon) at a final concentration of 5  $\mu$ g/ml. HA-tagged proteins were stained using monoclonal mouse HA antibodies (Sigma-Aldrich) at a dilution of 1:200. Alexa 488-, 561-, or 647-conjugated secondary antibodies (Invitrogen, Molecular Probes) were incubated in 3% BSA for 45 min at RT with a dilution of 1:1,000. Nuclei were stained using 4',6-diamidino-2-phenylindole (DAPI) for 1 min at a dilution of 1:4,000 after incubation with secondary antibodies. Cells were mounted with Fluoromount G (Southern Biotechnology Associates, Birmingham, AL, USA), and whole-cell z-stacks for PI4P analysis were acquired with a Leica SP2 confocal laser scanning microscope. Quantification of intracellular PI4P signals was described before (7). For colocalization analysis, samples were imaged with an UltraVIEW ERS spinning disk (PerkinElmer Life Sciences) on a Nikon TE2000-E inverted confocal microscope using a Plan-Apochromat VC 100 $\times$  objective (numeric aperture [NA], 1.4). Optical sections of 0.13  $\mu$ m were acquired separately for each channel. z-stacks were deconvolved with a theoretical point-spread function, and chromatic shifts between green and far-red dyes were corrected according to recordings of multifluorescent beads under equal conditions (Tetraspeck beads; Invitrogen) using Huygens Essential software (SVI, Netherlands). Three-dimensional (3D) colocalization of fluorescence signals was evaluated quantitatively for Pearson's correlation coefficient by using the integrated function in Huygens Essential.

**Metabolic radiolabeling of proteins, immunoprecipitation, and Western blotting.** Metabolic labeling and immunoprecipitation of NS5A JFH-1 or HA-PI4KIII $\alpha$  was described earlier (10). Polyclonal sheep antibodies directed against NS5A of JFH-1 were a generous gift of M. Harris (Leeds University, United Kingdom), and monoclonal mouse antibodies against HA tag were obtained from Sigma-Aldrich (H3663). For Western blotting, 1/10 of the cells in a T25 cell culture flask were denatured, heated in 2 $\times$  Laemmli buffer, and loaded onto an 8% polyacrylamide-SDS gel. After separation and transfer to a polyvinylidene difluoride (PVDF) membrane, PI4KIII $\alpha$  was detected using monoclonal rabbit antibodies (Cell Signaling). HA-tagged proteins or  $\beta$ -actin was detected using monoclonal mouse antibodies (H3663 or A5441, respectively; Sigma).

***In vitro* lipid kinase assay after immunoprecipitation of HA-tagged PI4KIII $\alpha$  deletion variants.** *In vitro* lipid kinase assays were performed as described previously by Berger et al. (33). Huh7-Lunet T7 shPI4K cells were seeded into a 10-cm dish and were transfected with HA-PI4KIII $\alpha$  deletion variants. Twenty-four hours posttransfection, cells were washed once with PBS and lysed in 1 ml lysis buffer (20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 2 mM EDTA, 1% NP-40, 10% glycerol, and protease inhibitor cocktail [Roche]) on ice for 1 h. After clearing of the lysate (14,000 rpm, 45 min, 4°C), proteins were immunoprecipitated overnight using agarose beads coupled to HA-specific antibodies (Sigma-Aldrich). Beads were washed twice with lysis buffer, once with PBS, and once with lipid kinase assay buffer (20 mM Tris-HCl, pH 7.5, 5 mM MgCl<sub>2</sub>, 0.5 mM EGTA, 0.4% Triton X-100). Beads were resuspended in 1 ml lipid kinase assay buffer, and 300  $\mu$ l was used for Western blot analysis. The remaining 700  $\mu$ l was spun down and resuspended in 35  $\mu$ l lipid kinase assay buffer. For measurement of *in vitro* lipid kinase activity, the following reaction mix was



prepared: 10  $\mu$ M phosphatidylinositol-phosphoserine (PI/PS; Invitrogen), 10  $\mu$ M ATP (Roche), 2 mM dithiothreitol (DTT), 10  $\mu$ Ci [ $\gamma$ - $^{32}$ P]ATP (Hartmann Analytik), and lipid kinase assay buffer in a final volume of 50  $\mu$ l. The reaction was started with the addition of 10  $\mu$ l bead slurry and incubated for 1 h at room temperature in a shaker. The reaction was stopped with the addition of 200  $\mu$ l 1 M HCl, and lipids were extracted by the addition of 400  $\mu$ l chloroform-methanol (1:1) suspension. After vortexing and centrifugation (13,000 rpm, 3 min), 125  $\mu$ l of the lower organic phase was transferred to a scintillation counting vial and 4 ml of Ultima Gold (PerkinElmer) was added. Incorporation of  $^{32}$ P was measured by using a liquid scintillation counter and is expressed as counts per minute (cpm).

**Statistical analysis and software.** Statistical analyses, as indicated in the figures, were performed using Microsoft Excel and GraphPad Prism software. Fluorescence images were analyzed using ImageJ/Fiji and Huygens Essential software. Figures were arranged with Adobe Photoshop and Adobe Illustrator.

**Homology modeling of PI4KIII $\alpha$   $\Delta$ 1-1535.** The delimitation of the PI3K-like region of PI4KIII $\alpha$  was obtained by a sequence alignment with PSI-BLAST (39) against the sequence of human PI4KIII $\alpha$ . Only PI3K elements with known structures were taken into account. In turn, a multiple alignment was generated with those results using COBALT (40). This led to a reliable alignment with structural templates matching PI4KIII $\alpha$  residues in the region 1536-2102. I-TASSER (41) was used to generate a structural model of PI4KIII $\alpha$   $\Delta$ 1-1535. The most relevant model was obtained with *Drosophila* PI3KIII VPS34 as the template (PDB code 2X6F) (42).

**Sequence-based domain predictions of PI4KIII $\alpha$   $\Delta$ 1536-2102.** We used the QUICK2D web service ([http://toolkit.tuebingen.mpg.de/quick2\\_d](http://toolkit.tuebingen.mpg.de/quick2_d)) from the MPI Bioinformatics Toolkit (43) to perform secondary structure and disorder analysis on the N terminus of the PI4KIII $\alpha$  sequence. QUICK2D runs parallel analyses with the most accurate secondary structure and disorder prediction programs. The reliability of the predictions was evaluated based on the agreement between different programs.

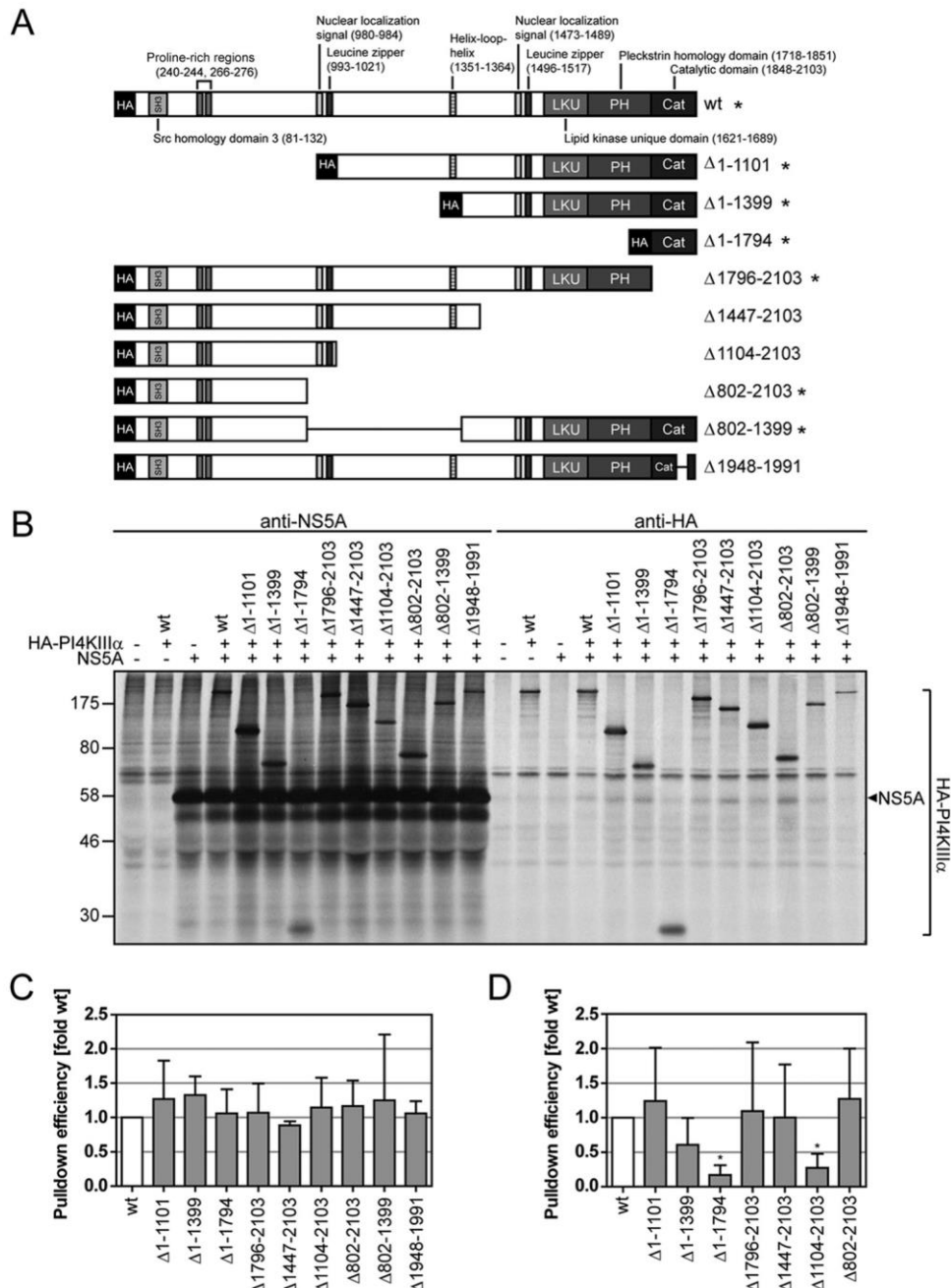
## RESULTS

**PI4KIII $\alpha$  interacts with NS5A via multiple interaction sites.** In this study, we aimed to investigate structural requirements of PI4KIII $\alpha$  enzymatic function and the requirements for HCV replication. Previously, we reported a crucial role of NS5A-PI4KIII $\alpha$  interaction in HCV replication by identifying NS5A residues important for PI4KIII $\alpha$  interaction and recruitment (10). Here, we first aimed to identify the binding site between NS5A and PI4KIII $\alpha$  by constructing different N- and C-terminal truncations of PI4KIII $\alpha$  that were tested for NS5A interaction by coimmunoprecipitation. We employed three serial deletions of the N terminus ( $\Delta$ 1-1101,  $\Delta$ 1-1399, and  $\Delta$ 1-1794), four serial deletions of the C terminus ( $\Delta$ 1796-2103,  $\Delta$ 1447-2103,  $\Delta$ 1104-2103, and  $\Delta$ 802-2103), and two internal deletions ( $\Delta$ 802-1399 and  $\Delta$ 1948-1991) encompassing functional motifs assigned in a previous study (Fig. 1A) (13). All mutants were expressed under transcriptional control of a T7 promoter and were cotransfected with a plasmid encoding NS5A of isolate JFH-1 into Huh7-Lunet cells stably expressing T7 RNA polymerase (Huh7-Lunet T7) followed by immunoprecipitation of NS5A or HA-tagged proteins. All PI4KIII $\alpha$  mutants were expressed in significant amounts, but very little NS5A coprecipitated in the HA IP (Fig. 1B, right). This was in line with previous results and likely was due to the much higher NS5A expression levels (10). However, when using NS5A-specific antibodies, all PI4KIII $\alpha$  mutants coprecipitated with NS5A (Fig. 1B, left, and C), and the specificity could not be increased by changing IP conditions (data not shown). Coprecipitation effi-

ciency was reduced for an NS5A variant harboring a mutation in the PI4KIII $\alpha$  interaction site (mutHIT) (10) but not to the same extent as the same mutant expressed in the context of the NS3 to NS5B polyprotein (termed NS3-5B) (10 and data not shown). Indeed, when NS5A was expressed in the context of the NS3-5B polyprotein, coprecipitation seemed to be more specific and the efficiency was significantly reduced for two deletion mutants ( $\Delta$ 1-1794 and  $\Delta$ 1104-2103, respectively) (Fig. 1D and data not shown). Based on these data, several regions within PI4KIII $\alpha$  still seemed to independently contribute to the interaction with NS5A. However, we regarded it unlikely that all of them would contribute to the function of PI4KIII $\alpha$  in HCV replication, since we had been able to assign a single 7-residue major functional PI4KIII $\alpha$  interaction motif in NS5A (10). Therefore, we sought alternative ways to characterize regions of PI4KIII $\alpha$  that are functionally important for HCV replication.

**Deletion of the N-terminal 1,101 amino acids of PI4KIII $\alpha$  has no effect on HCV replication.** To identify regions within PI4KIII $\alpha$  important for HCV, we analyzed a set of PI4KIII $\alpha$  mutants for their ability to support HCV replication (Fig. 1A, marked with asterisks), including a full-length PI4KIII $\alpha$  variant carrying a point mutation at position 1957, thereby rendering the protein enzymatically inactive (D1957A) (7). HA-tagged shRNA-resistant PI4KIII $\alpha$  variants were expressed in Huh7-Lunet cells with a constitutive knockdown of endogenous PI4KIII $\alpha$  (shPI4K) using selectable lentiviral vectors (Fig. 2B). Endogenous PI4KIII $\alpha$  was found in very small amounts in the control cell line expressing a nontargeting shRNA (NT) (Fig. 2B, black arrowhead) and was absent from the shPI4K cell line. All PI4KIII $\alpha$  mutants were expressed at higher levels than endogenous PI4KIII $\alpha$  in the NT cell line. Subgenomic JFH-1 firefly luciferase reporter replicons were used to analyze the ability of PI4KIII $\alpha$  mutants to support HCV replication (Fig. 2A). Knockdown of endogenous PI4KIII $\alpha$  resulted in an approximately 100-fold reduction of HCV replication compared to the NT cell line at all measured time points, which was rescued by expression of PI4KIII $\alpha$  wild type (wt) but not of the inactive D1957A mutant, as expected (Fig. 2C). Interestingly, HCV replication efficiency was not enhanced beyond the level of the control cells, despite strongly enhanced PI4KIII $\alpha$  expression levels in cells ectopically expressing the full-length enzyme (Fig. 2B, lower, NT compared to wt), suggesting that small amounts of active PI4KIII $\alpha$  already were sufficient to achieve maximal HCV replication efficiency. Remarkably, a cell line expressing a PI4KIII $\alpha$  mutant lacking the N-terminal 1,101 amino acids ( $\Delta$ 1-1101) still was able to support HCV replication to the same extent as the NT or PI4KIII $\alpha$  wt cell line. All other PI4KIII $\alpha$  variants were not able to rescue HCV replication. Still, a limited increase in HCV replication compared to the shPI4K or D1957A cell line was observed for some of the mutants, including C-terminal deletions lacking the catalytic center of the lipid kinase. This might argue for an indirect role of other structural features of PI4KIII $\alpha$  in HCV replication or for subtle differences in PI4KIII $\alpha$  knockdown efficiency in those cell lines. Together, these findings demonstrated that deletion of the N-terminal 1,101 amino acids did not compromise the functions of PI4KIII $\alpha$  in HCV replication, whereas further truncations of the N terminus or all deletions in the C-terminal half of PI4KIII $\alpha$  were deleterious.

**Deletion of the N-terminal 1,101 amino acids has no effect on PI4KIII $\alpha$  enzymatic activity.** Our previous studies had shown that HCV replication or expression of HCV NS3 to NS5B acti-

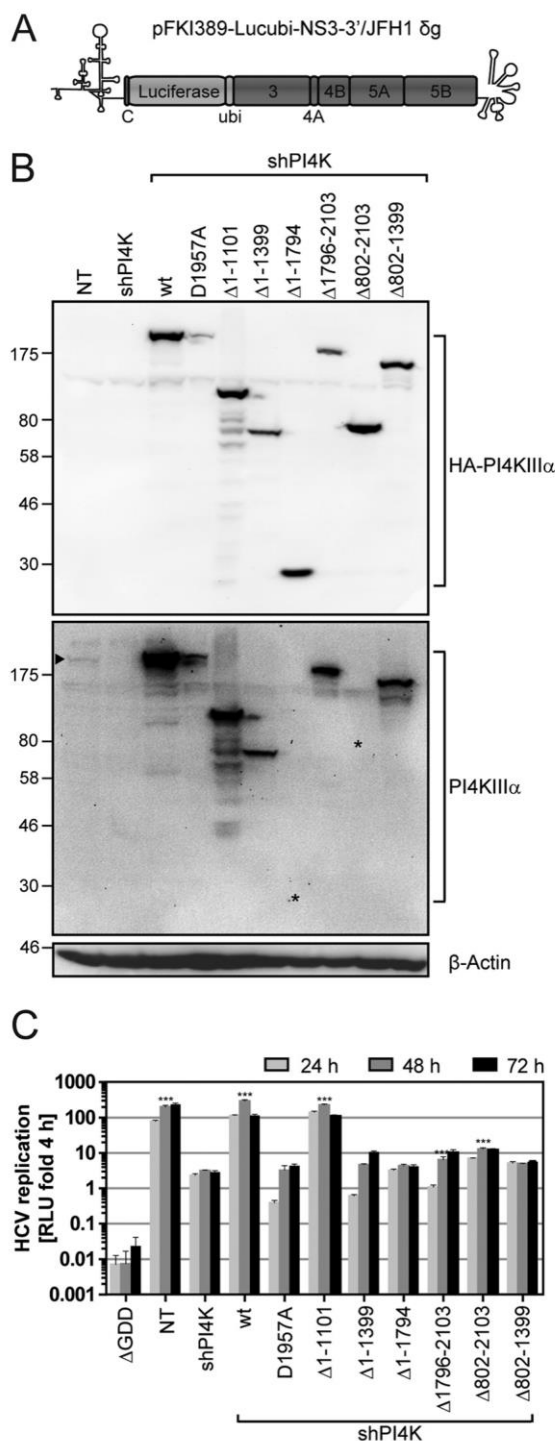


**FIG 1** PI4KIII $\alpha$  interacts with NS5A via multiple interaction sites. (A) Schematic representation of HA-tagged, truncated PI4KIII $\alpha$  proteins used for NS5A interaction mapping and subsequent analyses. Annotations of subdomains and motifs were adapted from Gehrmann et al. (13) and are indicated with the respective amino acid positions. Residues encompassing the respective truncations are indicated at the right. Asterisks denote mutants used for further functional characterization. All numberings refer to GenBank accession number [AB384703](https://www.ncbi.nlm.nih.gov/nuclot/AB384703). (B, C, and D) Plasmids encoding different HA-PI4KIII $\alpha$  truncations were cotransfected with constructs encoding HCV NS5A (B and C) or NS3-5B (D) into Huh7-Lunet T7 cells. Proteins were metabolically labeled with [ $^{35}$ S]methionine/cysteine and immunoprecipitated using polyclonal antibodies directed against NS5A or monoclonal antibodies directed against the HA peptide, respectively. Proteins were separated by 10% SDS-PAGE and visualized by autoradiography. (C and D) Bands obtained from experiments shown in panel B were quantified by densitometry. Pull-down efficiency of PI4KIII $\alpha$  mutants by NS5A is shown as a ratio of total amounts of HA-PI4KIII $\alpha$  to coprecipitated protein by NS5A. Shown are mean values and standard deviations from three different experiments.

vated lipid kinase activity of PI4KIII $\alpha$ , resulting in substantially increased amounts of intracellular PI4P (7, 10). Therefore, increased intracellular PI4P levels in the presence of HCV can be used as an indirect measure of intracellular PI4KIII $\alpha$  activity and for a functional interaction with HCV proteins. Thus, we charac-

terized the same set of shPI4K Lunet-T7 cell lines stably expressing the PI4KIII $\alpha$  mutants for the induction of PI4P in the presence of HCV. Since most of the mutants did not support HCV replication (Fig. 2C), we ectopically expressed the NS3-5B polyprotein under transcriptional control of the T7 promoter. In mock-transfected





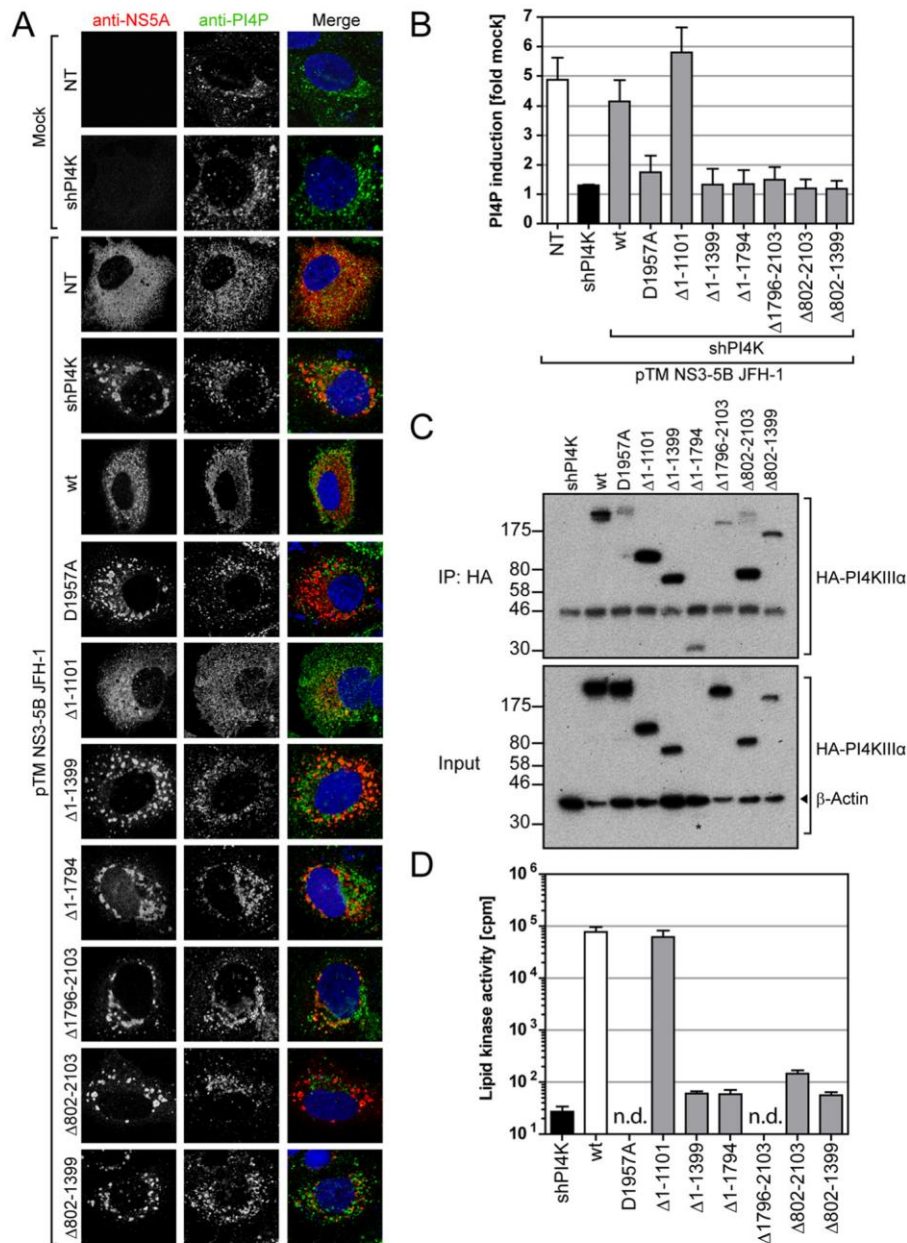
**FIG 2** Deletion of the N-terminal 1,101 amino acids of PI4KIIIα does not compromise its function in HCV replication. (A) Schematic representation of the subgenomic firefly luciferase reporter replicon of HCV genotype 2a (isolate JFH-1) used in this study. (B) Huh7-Lunet T7 cells were stably transduced with selectable lentiviral vectors expressing nontargeting shRNA (NT) or shRNA targeting endogenous PI4KIIIα (shPI4K) and encoding HA-tagged PI4KIIIα variants as indicated. Knockdown of endogenous PI4KIIIα and ectopic expression of HA-tagged PI4KIIIα variants was assessed by Western blotting using monoclonal antibodies against the HA peptide (top) or polyclonal antibodies against PI4KIIIα (middle), respectively. Asterisks indicate PI4KIIIα variants which have lost epitopes recognized by the polyclonal antiserum. An arrowhead points to endogenous PI4KIIIα. (C) Replication of HCV was quan-

NT cells, PI4P localized mainly to the Golgi compartment in a perinuclear area, as reported previously (Fig. 3A) (7, 10, 23). Knockdown of PI4KIIIα did not have any effects on the distribution or the amount of PI4P, as expected, since these PI4P pools are synthesized mainly by the related PI4KIIIβ isoform (16) (Fig. 3A, mock shPI4K). In NT cells expressing HCV NS3 to NS5B, PI4P was found in large amounts at intracellular membranes all over the cell, only partially colocalizing with HCV NS5A (Fig. 3A), and total PI4P amounts were 4- to 5-fold increased compared to levels for mock-transfected cells (Fig. 3B). Knockdown of PI4KIIIα abolished the HCV-dependent induction of these PI4P pools (Fig. 3B) and also altered the subcellular distribution of NS5A to a more clustered morphology (Fig. 3A, pTM NS3-5B JFH-1, shPI4K) (7). Both phenotypes, i.e., induction of PI4P as well as localization of NS5A, were rescued in cell lines expressing PI4KIIIα wt (Fig. 3A and B). Among the PI4KIIIα deletion mutants, only cell lines harboring Δ1-1101 exhibited high intracellular PI4P pools and NS5A distribution comparable to that of the NT cell line and to PI4KIIIα wt expression, suggesting that the N-terminal 1,101 amino acids of PI4KIIIα were dispensable for the activation of PI4KIIIα by the HCV nonstructural proteins. To address the impact of the deletions on enzymatic activity of PI4KIIIα more directly, we analyzed whether these mutants exhibited lipid kinase activity *in vitro*. HA-tagged PI4KIIIα protein variants were transiently expressed under transcriptional control of the T7 promoter in shPI4K cells constitutively expressing T7 RNA polymerase to exclude any cross-contamination with endogenous PI4KIIIα and were purified by immunoprecipitation using HA-specific antibodies (Fig. 3C, upper) and subjected to an *in vitro* lipid kinase assay. Expression levels (Fig. 3C, lower) and purification efficiency varied to some extent among different constructs; still, in this experiment, all PI4KIIIα variants except D1957A and Δ1796-2103 could be recovered in sufficient amounts (Fig. 3C, upper). Again, only PI4KIIIα wt and the Δ1-1101 mutant showed *in vitro* lipid kinase activity above background levels (Fig. 3D).

Collectively, our results indicated that the N-terminal 1,101 amino acids of PI4KIIIα were not required for enzymatic activity of PI4KIIIα or for functional interaction with HCV, whereas larger N-terminal deletions (Δ1-1399) or deletions within the C terminus were not tolerated.

**Subcellular localization of HA-PI4KIIIα mutants and recruitment by NS5A.** PI4KIIIα has been described to contribute primarily to PI4P synthesis in the ER and the plasma membrane, in contrast to other PI4K isoforms (reviewed in reference 16). Therefore, we next analyzed the subcellular localization of the PI4KIIIα deletion mutants to identify determinants of subcellular localization by transient transfection of T7-driven constructs (Fig. 4A). Interestingly, PI4KIIIα wt as well as all mutants were distributed broadly throughout the cytoplasm of the transfected cells and did not significantly colocalize with the ER marker protein disulfide isomerase (PDI) or a Golgi marker (Fig. 4A and data not shown), in line with a recent report (22). Only the PI4KIIIα vari-

ties by transfection of firefly luciferase reporter replicons into the indicated cell lines. Replication efficiency is expressed as relative light units (RLU) normalized to 4 h after transfection to account for transfection efficiency. A replication-deficient deletion mutant in NS5B (ΔGDD) served as the negative control. Significance above values for shPI4K cells is shown for replication efficiencies 48 h after transfection. Data represent mean values and standard deviations from a representative experiment ( $n = 3$ ). \*\*\*,  $P < 0.001$ .



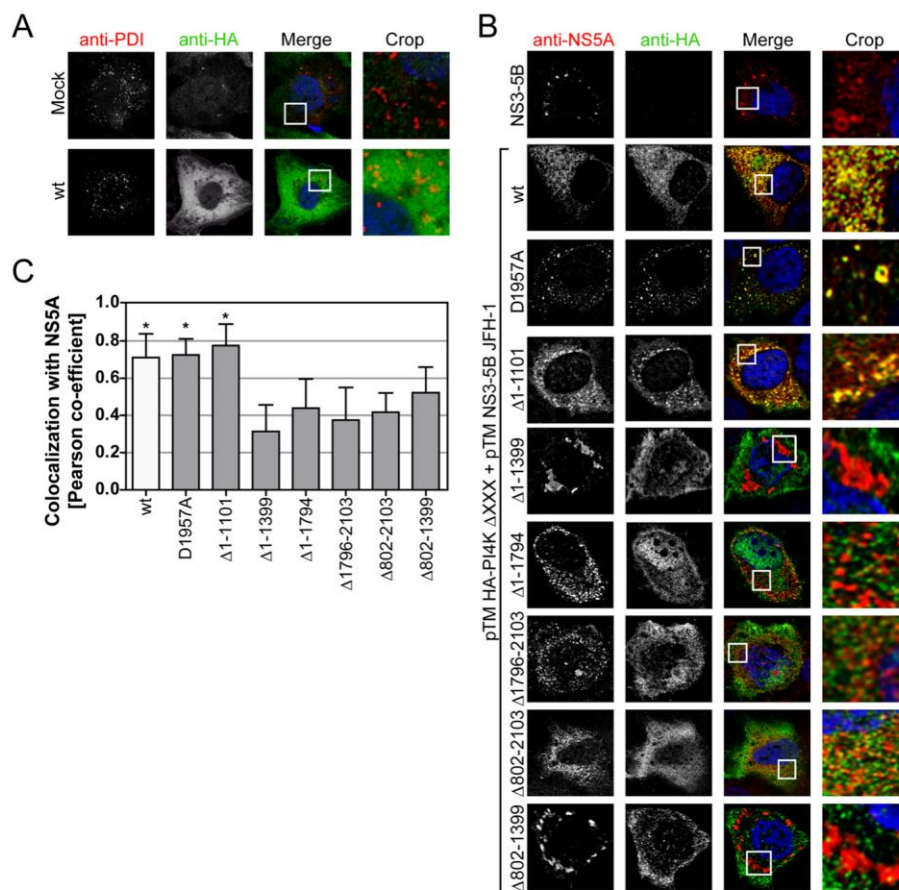
**FIG 3** Lipid kinase activity of truncated PI4KIIIα proteins *in vitro* and activation by HCV in cell culture. (A) Plasmids encoding HCV NS3-5B were transfected into Huh7-Lunet T7 cell lines depleted of endogenous PI4KIIIα and stably expressing the indicated variants of PI4KIIIα (for details, refer to the legend of Fig. 2). Twenty-four h after transfection, cells were fixed and permeabilized with digitonin, and NS5A (red) and PI4P (green) were stained using the respective monoclonal antibodies. Nuclei (blue) were stained using DAPI. (B) Fluorescence signals of PI4P from whole-cell z-stacks were quantified using ImageJ software. PI4P induction was quantified by normalizing intensities of HCV-transfected cells to those of mock-transfected cells of each cell line. Standard deviations are indicated. At least 30 cells were quantified for each condition. (C) pTM constructs encoding the indicated HA-tagged PI4KIIIα mutants were transfected into Huh7-Lunet T7 shPI4K cells. Twenty-four h after transfection, cells were lysed and HA-tagged proteins were immunoprecipitated using HA beads. PI4KIIIα and β-actin were visualized by Western blotting, using antibodies against the HA peptide and β-actin, respectively, in the cell lysate (Input) and after IP. An asterisk indicates the position of a weak band of the Δ1-1794 mutant. The ca. 46-kDa band in the IP fractions corresponds to the heavy chain of the antibodies used for IP. (D) Immunoprecipitated HA-PI4KIIIα mutants were subjected to *in vitro* lipid kinase assays based on incorporation of <sup>32</sup>P into phosphatidylinositol substrates. Radioactive lipids were extracted and subjected to liquid scintillation counting. Depicted are counts per minute (cpm) obtained for the different PI4KIIIα mutants after background subtraction. n.d., not determined due to low recovery levels. The lack of enzymatic activity of mutants D1957A and Δ1796-2103 was determined in independent experiments (Fig. 5E and data not shown, respectively).

ant with the largest N-terminal deletion (Δ1-1794) had a more prominent nuclear localization (data not shown), which might be due either to a loss of a nuclear export signal or, more likely, because of the small size of this variant (ca. 30 kDa), allowing diffusion through the nuclear pore (44).

In essence, although we have generated a set of huge, nonoverlapping deletions of PI4KIIIα, all mutants showed almost the same broad subcellular distribution and no clear colocalization to particular intracellular membranes.

We have shown before that PI4KIIIα is recruited to the MW,





**FIG 4** Subcellular localization of PI4KIII $\alpha$  truncation mutants in the presence and absence of HCV NS proteins. Huh7-Lunet shPI4K cells were transfected with plasmids encoding the indicated HA-tagged PI4KIII $\alpha$  variants (A, B, and C) and with constructs expressing HCV NS3 to NS5B (B and C). (A) Proteins were stained using monoclonal antibodies directed against PDI as the marker of the ER (red) or HA peptide (green), respectively. Nuclei (blue) were stained using DAPI. (B) HCV NS5A or HA-tagged proteins were stained using monoclonal antibodies directed against NS5A (red) or HA peptide (green), respectively. (C) 3D colocalization of NS5A and PI4KIII $\alpha$  in experiments as shown in panel B was assessed by recording whole-cell z-stacks and analyzed using Huygens Essential software. Depicted are mean values and standard deviations of Pearson correlation coefficients from at least 10 individual cells per condition. \*,  $P < 0.05$ .

which might be crucial for HCV replication (7). Therefore, we coexpressed NS3 to NS5B with the different HA-tagged PI4KIII $\alpha$  mutants and found that PI4KIII $\alpha$  wt and the inactive D1957A point mutant widely colocalized with NS5A in distinct foci throughout the cell, arguing for a specific recruitment of PI4KIII $\alpha$  independent of enzymatic activity. Importantly, only mutant  $\Delta$ 1-1101 colocalized to NS5A-positive compartments in the same way as PI4KIII $\alpha$  wt, whereas all other deletion mutants showed no specific recruitment to NS5A at all (Fig. 4B and C). These data indicate that recruitment of PI4KIII $\alpha$  to NS5A-positive compartments seemed not to depend on a distinct interaction region but rather relied on an overall structurally intact kinase.

In summary, gross mapping of functionally important regions of PI4KIII $\alpha$  revealed that the N-terminal 1,101 amino acids were dispensable for enzymatic activity, subcellular localization, specific recruitment to the HCV replication sites, and HCV replication. In contrast, mutants with further N-terminal deletions and all C-terminal deletions had lost enzymatic activity and recruitment by HCV and had failed to support HCV replication (Table 1).

**N-terminal deletions of up to 1,152 amino acids do not compromise major PI4KIII $\alpha$  functions.** Our previous experiments clearly revealed that the region between residues 1102 and 1399

was critical for the enzymatic function of PI4KIII $\alpha$  as well as for its role in HCV replication. Therefore, we next aimed to define more precisely the N-terminal boundary critical for PI4KIII $\alpha$  function. We constructed deletion mutants lacking the N-terminal 1,152, 1,202, or 1,252 amino acids, respectively, and stably (Fig. 5A to C) or transiently (Fig. 5D to F) expressed these mutants in shPI4K cells. We analyzed these cell lines for their ability to support HCV replication (Fig. 5B), for the activation of enzymatic activity by HCV NS3 to NS5B (Fig. 5C), and for enzymatic activity *in vitro* (Fig. 5D and E). All functions were fully retained by mutant  $\Delta$ 1-1152, whereas mutants  $\Delta$ 1-1202 and  $\Delta$ 1-1252 showed dramatically reduced or no rescue of HCV replication, respectively (Fig. 5B), no significant induction of intracellular PI4P levels upon expression of NS3 to NS5B (Fig. 5C), and no significant lipid kinase activity *in vitro* (Fig. 5E), as also observed for other enzymatically inactive mutants (compare Fig. 5 to Fig. 2 and 3 as well as Table 1). Although mutant  $\Delta$ 1-1202 was devoid of any detectable enzymatic activity *in vitro*, it still rescued HCV replication and intracellular PI4P induction to some extent, arguing for a minimal residual activity that is sufficient for limited HCV replication. Subcellular localization of all three mutants was not distinguishable from PI4KIII $\alpha$  wt in the absence of HCV (data not shown),

TABLE 1 Summary of phenotypes observed for the different PI4KIII $\alpha$  mutant variants

PI4KIII $\alpha$ type	NS5A interaction <sup>a</sup>	NS5A interaction <sup>a</sup> (NS3-5B)	HCV replication <sup>b</sup>	PI4P induction <sup>c</sup> and <i>in vitro</i> lipid kinase activity	Recruitment to membranous web <sup>d</sup>
wt	+	+	+	+	+
D1957A	NA	+ <sup>e</sup>	—	—	+
$\Delta$ 1-1101	+	+	+	+	+
$\Delta$ 1-1152	NA	NA	+	+	+/-
$\Delta$ 1-1202	NA	NA	—	—	—
$\Delta$ 1-1252	NA	NA	—	—	—
$\Delta$ 1-1399	+	+	—	—	—
$\Delta$ 1-1794	+	—	—	—	—
$\Delta$ 1796-2103	+	+	—	—	—
$\Delta$ 802-2103	+	+	—	—	—
$\Delta$ 802-1399	+	NA	—	—	—
$\Delta$ 1-1101 $\Delta$ DII	NA	NA	—	—	—
$\Delta$ 1-1101 $\Delta$ DIII	NA	NA	—	—	—
$\Delta$ 1-1101 $\Delta$ HLLH	NA	NA	—	—	—
$\Delta$ 1-1101 $\Delta$ NLS	NA	NA	+	+	+/-
$\Delta$ 1-1101 $\Delta$ LZ	NA	NA	—	—	—
$\Delta$ 1-1101 $\Delta$ LKU	NA	NA	—	—	+/-
$\Delta$ 1-1101 $\Delta$ PH	NA	NA	—	—	+/-

<sup>a</sup> A plus sign indicates the NS5A interaction is not significantly impaired compared to that of PI4KIII $\alpha$  wt (Fig. 1). NA, not analyzed.

<sup>b</sup> A plus sign indicates the rescue of HCV replication is not significantly impaired compared to that of the NT cell line (Fig. 2, 5, and 6).

<sup>c</sup> For PI4P induction, a plus sign indicates intracellular PI4P is significantly increased compared to that of mock-transfected cells (Fig. 3 and 7). For *in vitro* lipid kinase activity, a plus sign indicates *in vitro* lipid kinase activity is at least 10% of that of the corresponding positive control (Fig. 3 and 7).

<sup>d</sup> A plus sign indicates that the Pearson correlation is significant over the number of analyzed samples (Fig. 4, 5, and 8). +/-, mutants do not show significant Pearson correlation values but showed substantially higher correlation than other mutants in the respective experiments.

<sup>e</sup> Adapted from reference 10.

but recruitment to NS5A-positive compartments was completely lost for mutants  $\Delta$ 1-1202 and  $\Delta$ 1-1252, whereas colocalization of mutant  $\Delta$ 1-1152 with NS5A-positive dots was slightly but significantly reduced compared to mutant  $\Delta$ 1-1101 (Fig. 5F), suggesting that region 1102-1202 plays a role in specific recruitment of PI4KIII $\alpha$  to the HCV replication sites.

In conclusion, we found that the N-terminal 1,152 amino acids of PI4KIII $\alpha$  are not required for lipid kinase activity or HCV replication.

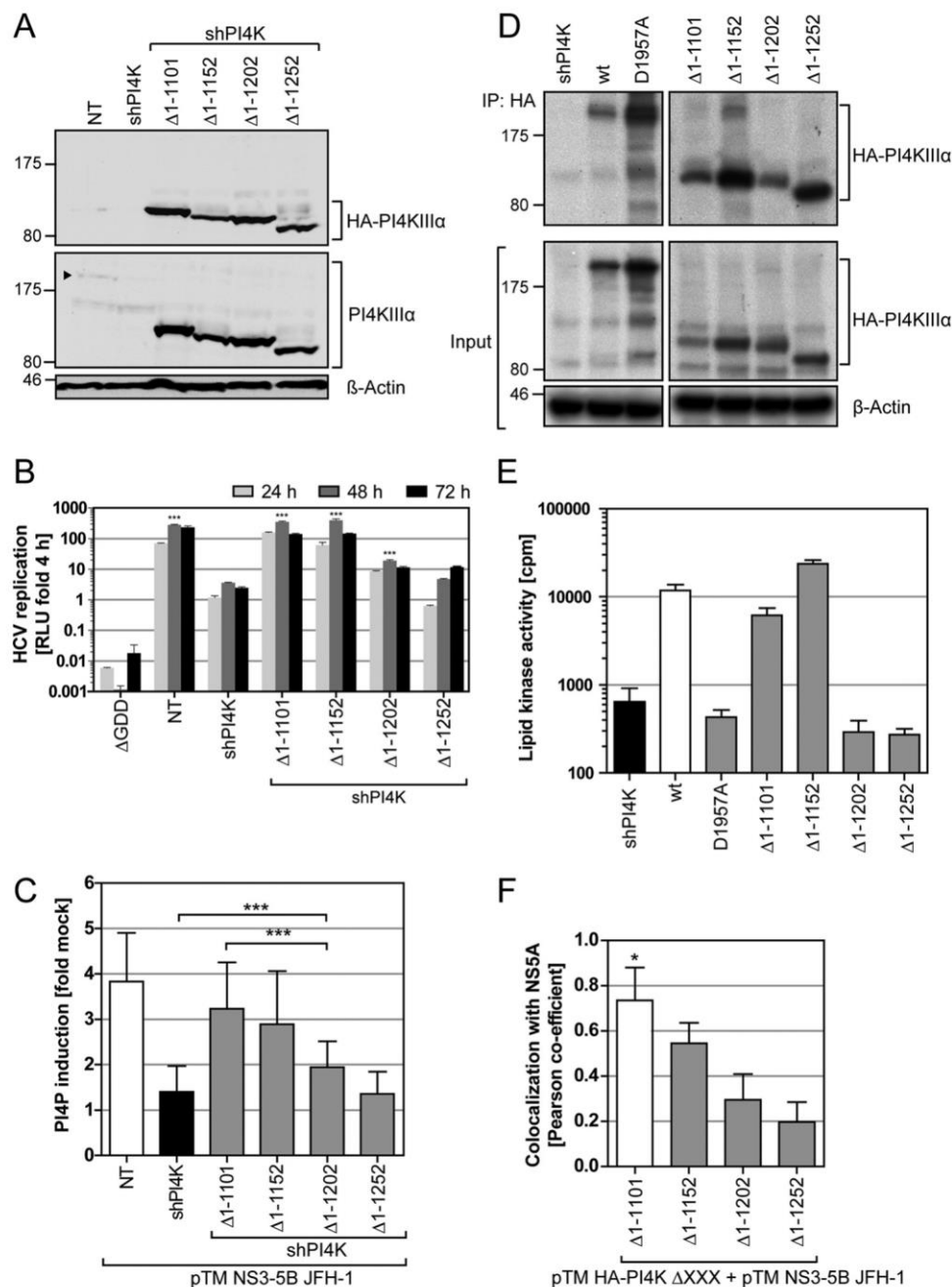
**Functional analysis of subdomains and predicted motifs in the C-terminal region of PI4KIII $\alpha$ .** After definition of the N-terminal boundaries, we next focused on C-terminal motifs of PI4KIII $\alpha$  and their contributions to HCV replication and enzymatic activity. Based on the domain annotations by Gehrmann et al. (13), we individually deleted the nuclear localization signal ( $\Delta$ NLS), the helix-loop-helix motif ( $\Delta$ HLLH), the leucine zipper ( $\Delta$ LZ), the lipid kinase unique domain ( $\Delta$ LKU), or the PH domain ( $\Delta$ PH) (Fig. 1A and 6A). We additionally included deletions encompassing residues 1280 to 1430 (designated DII) or 1430 to 1530 (designated DIII) (see Fig. 9A), respectively, since these regions were predicted to fold into distinct alpha-helical subdomains (see Fig. 9A and B). We chose the N-terminally truncated mutant  $\Delta$ 1-1101 with an N-terminal HA tag as the backbone for all further deletions of C-terminal domains (Fig. 6A), since this mutant was identical to the full-length enzyme regarding recruitment by HCV, rescue of HCV replication, and induction of PI4P synthesis by HCV and was similar to the full-length enzyme regarding enzymatic activity. Using this truncated variant facilitated expression and detection of the mutants, although we cannot fully exclude that some deletion mutants might behave differently in the context of the full-length enzyme.

To address their ability to rescue HCV replication, mutants

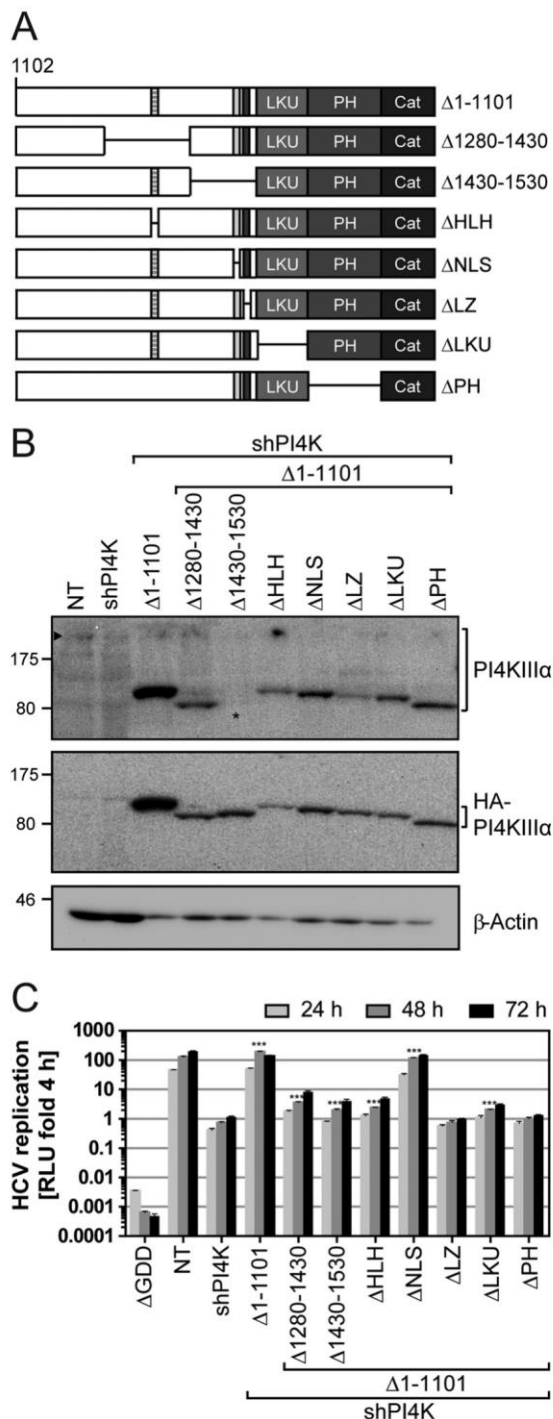
were expressed in shPI4K cells by lentiviral transfer (Fig. 6B), and HCV replication was assessed as before by transfection of subgenomic JFH-1 luciferase reporter replicons using the cell line expressing  $\Delta$ 1-1101 as a positive control (Fig. 6C). Only the mutant lacking the proposed NLS was still able to fully support HCV replication comparable to that of the  $\Delta$ 1-1101 mutant or the NT cell line, whereas expression of all other mutants did not substantially rescue HCV replication, apart from some minor increases similar to those observed before (compare Fig. 6C to 2C). Thus, the region encompassing the NLS seemed to be dispensable for HCV replication. Similar results were obtained by analysis of the ability of the mutants to give rise to increased PI4P levels upon expression of HCV NS3 to NS5B (Fig. 7A and B) and lipid kinase activity *in vitro* (Fig. 7C and D). Again, all mutants except  $\Delta$ NLS were inactive. However, stimulation of PI4P synthesis in cell culture (Fig. 7A and B) as well as lipid kinase activity (Fig. 7C and D) of  $\Delta$ NLS was strongly reduced compared to that of the parental  $\Delta$ 1-1101 mutant. These findings demonstrate that deletion of the NLS partly compromised enzymatic activity of PI4KIII $\alpha$ , which, however, still was sufficient to fully support HCV replication.

We next examined whether the deletions affect intracellular distribution of PI4KIII $\alpha$ . As observed for the larger deletions described above, all mutants again localized broadly throughout the cytoplasm in the absence of HCV proteins, similar to the parental  $\Delta$ 1-1101 mutant (Fig. 8A and data not shown). Hence, we were not able to identify regions that are solely responsible for correct localization of PI4KIII $\alpha$ . In contrast, all deletions affected colocalization with NS5A upon coexpression of NS3 to NS5B, albeit to various extents (Fig. 8B and C). Deletion of residues 1280 to 1430 (DII) or 1430 to 1530 (DIII) in the context of  $\Delta$ 1-1101 resulted in a complete loss of NS5A colocalization (Fig. 8B and C), arguing for an important role of these regions in recruitment of PI4KIII $\alpha$





**FIG 5** Mapping of the N-terminal border required for PI4KIII $\alpha$  function. (A) Huh7-Lunet T7 cells were stably transduced with selectable lentiviral vectors expressing nontargeting shRNA (NT) or shRNA targeting endogenous PI4KIII $\alpha$  (shPI4K) and encoding HA-tagged PI4KIII $\alpha$  variants as indicated. The knockdown of endogenous PI4KIII $\alpha$  and ectopic expression of HA-tagged PI4KIII $\alpha$  variants was assessed by Western blotting using polyclonal antibodies against PI4KIII $\alpha$  (middle) or monoclonal antibodies against the HA peptide (top). An arrowhead points to endogenous PI4KIII $\alpha$ . (B) Replication of HCV was quantified by transfection of firefly luciferase reporter replicons into the indicated cell lines. Replication efficiency is expressed as relative light units (RLU) normalized to 4 h after transfection to account for transfection efficiency. A replication-deficient deletion mutant in NS5B ( $\Delta$ GDD) served as the negative control. Data represent mean values and standard deviations from a representative experiment ( $n = 3$ ). Significance over values for shPI4K cells is shown for replication efficiencies 48 h after transfection. (C) Plasmids encoding HCV NS3-5B were transfected into Huh7-Lunet T7 cell lines depleted of endogenous PI4KIII $\alpha$  and stably expressing the indicated variants of PI4KIII $\alpha$  (for details, refer to the legend of Fig. 2). Twenty-four h after transfection, cells were stained for PI4P and fluorescent signals were quantified by normalizing intensities of HCV-transfected cells to mock-transfected cells of each cell line. Standard deviations are indicated. At least 30 cells were quantified for each condition. (D) pTM constructs encoding the indicated HA-tagged PI4KIII $\alpha$  mutants were transfected into Huh7-Lunet T7 shPI4K cells. Twenty-four h after transfection, cells were lysed and HA-tagged proteins were immunoprecipitated using HA beads. Input and immunoprecipitated proteins were visualized by monoclonal antibodies directed against HA peptide or  $\beta$ -actin, respectively. Shown are cropped images from the same Western blot. (E) Immunoprecipitated HA-PI4KIII $\alpha$  mutants were subjected to *in vitro* lipid kinase assays based on incorporation of  $^{32}$ P into phosphatidylinositol substrates. Radioactive lipids were extracted and subjected to liquid scintillation counting. Depicted are counts per minute (cpm) obtained for the different PI4KIII $\alpha$  mutants. (F) Huh7-Lunet shPI4K cells were cotransfected with plasmids encoding the indicated HA-tagged PI4KIII $\alpha$  variants and with constructs expressing HCV NS3 to NS5B. HCV NS5A- or HA-tagged proteins were stained using monoclonal antibodies. Depicted are mean values and standard deviations of Pearson correlation coefficients from at least 10 individual cells per condition. \*\*\*,  $P < 0.001$ ; \*,  $P < 0.05$ .



**FIG 6** Importance of individual PI4KIIIα subdomains for HCV replication. (A) Schematic representation of the HA-tagged, C-terminal half of the PI4KIIIα protein used for further mapping of functional domains. Annotations of subdomains and motifs were adapted from Gehrmann et al. (13) or based on structure modeling (Δ1280-1430 and Δ1430-1530; for details, refer to the legends to Fig. 1 and 9, respectively). All numberings refer to GenBank accession number AB384703. (B) Huh7-Lunet T7 cell lines were stably transduced with selectable lentiviral vectors expressing nontargeting shRNA (NT) or shRNA targeting endogenous PI4KIIIα (shPI4K) and encoding HA-tagged PI4KIIIα variants as indicated. Knockdown of endogenous PI4KIIIα and ectopic expression of HA-tagged PI4KIIIα variants was verified by Western blotting using polyclonal antibodies against PI4KIIIα (top) or monoclonal antibodies against the HA peptide (middle). An arrowhead points to endogenous

by HCV or for the structural integrity of the enzyme. Deletions of smaller motifs within DII and DIII (ΔNLS, ΔHLH, and ΔLZ) affected NS5A colocalization to various extents, probably reflecting the impairment of the overall structure rather than targeting a specific interaction motif with HCV proteins. However, mutants ΔLKU and ΔPH still partly colocalized with NS5A despite their lack of enzymatic activity, suggesting that these regions are not essential for recruitment by the HCV replicase.

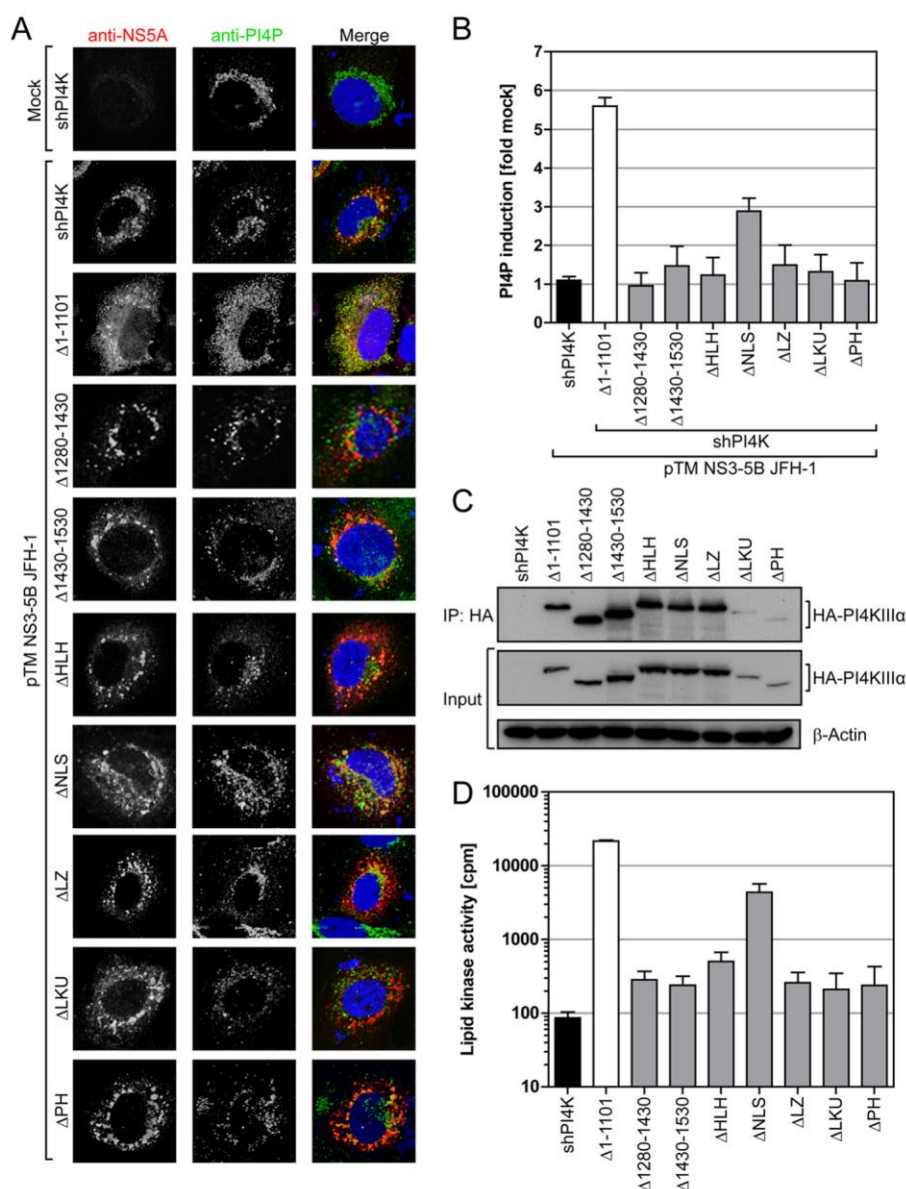
Taken together, our functional studies on the C-terminal half of PI4KIIIα demonstrated a tight linkage between rescue of HCV replication, PI4P induction, and lipid kinase activity *in vitro* (Table 1). All deletions analyzed in the context of Δ1-1101 except ΔNLS entirely abrogated enzymatic activity, suggesting that the entire C-terminal half of PI4KIIIα (aa 1152 to 2103) is involved in generating the catalytically active enzyme and that any gross deletion in this region is deleterious for structural integrity. Recruitment of PI4KIIIα by the HCV proteins was not entirely abrogated by some of the deletions, suggesting that this process was mediated by different structural elements at the C-terminal half.

**Homology modeling of the C terminus of PI4KIIIα allows refinement of domain annotations.** Structural information of domains at the C-terminal part of PI4KIIIα was not available, but these domains seemed to play important roles in enzymatic function and recruitment by HCV. Therefore, we used secondary structure predictions and structural modeling of residues 1102 to 2102 to identify motifs and domains which might be involved in those functions (summarized in Fig. 9A). The region encompassing residues 1101 to 1535 has no detectable homology to any protein of known structure, even using the most sensitive detection and structure prediction algorithms available (45). However, based on secondary structure prediction, it is likely almost all helical. Furthermore, there are clear breaks in the predicted secondary structure associated with a strong propensity to disorder in the sequences flanking residues 1130 to 1270 (Fig. 9B). This suggests that there is a distinct α-helical domain or subdomain in this region, which we termed DI. Breaks are less obvious in the stretch of residues 1271 to 1535, but there may be two more largely α-helical (sub)domains, DII and DIII, further downstream (not shown).

For modeling the very C terminus of PI4KIIIα, we used structural information of phosphoinositide 3-kinases (PI3K), which are distant homologs of type III phosphoinositide 4-kinases (46), a relationship that is reflected by their common sensitivity to wortmannin. Shortly after Gehrmann et al. identified and cloned human PI4KIIIα (47), the first crystal structures of PI3Ks became available (48). Using PI3Ks as templates, it is possible to accurately model the C-terminal part of PI4KIIIα for residues 1536 to 2102, with the exception of residues 1810 to 1840, which constitute an insertion in PI4KIIIα compared to PI3K (Fig. 9C, marked with an asterisk). As for PI3K (48), the PI4KIIIα 1536-2102 model is best

PI4KIIIα, and an asterisk points to a PI4KIIIα variant which was not recognized by the polyclonal antiserum. (C) Replication of HCV was quantified by transfection of firefly luciferase reporter replicons into the indicated cell lines. Replication efficiency is expressed as relative light units (RLU) normalized to 4 h after transfection for transfection efficiency. A replication-deficient deletion mutant in NS5B (ΔGDD) served as a negative control. Significance over values for shPI4K cells are shown for replication efficiencies 48 h after transfection. Data represent mean values and standard deviations from a representative experiment ( $n = 3$ ). \*\*\*,  $P < 0.001$ .

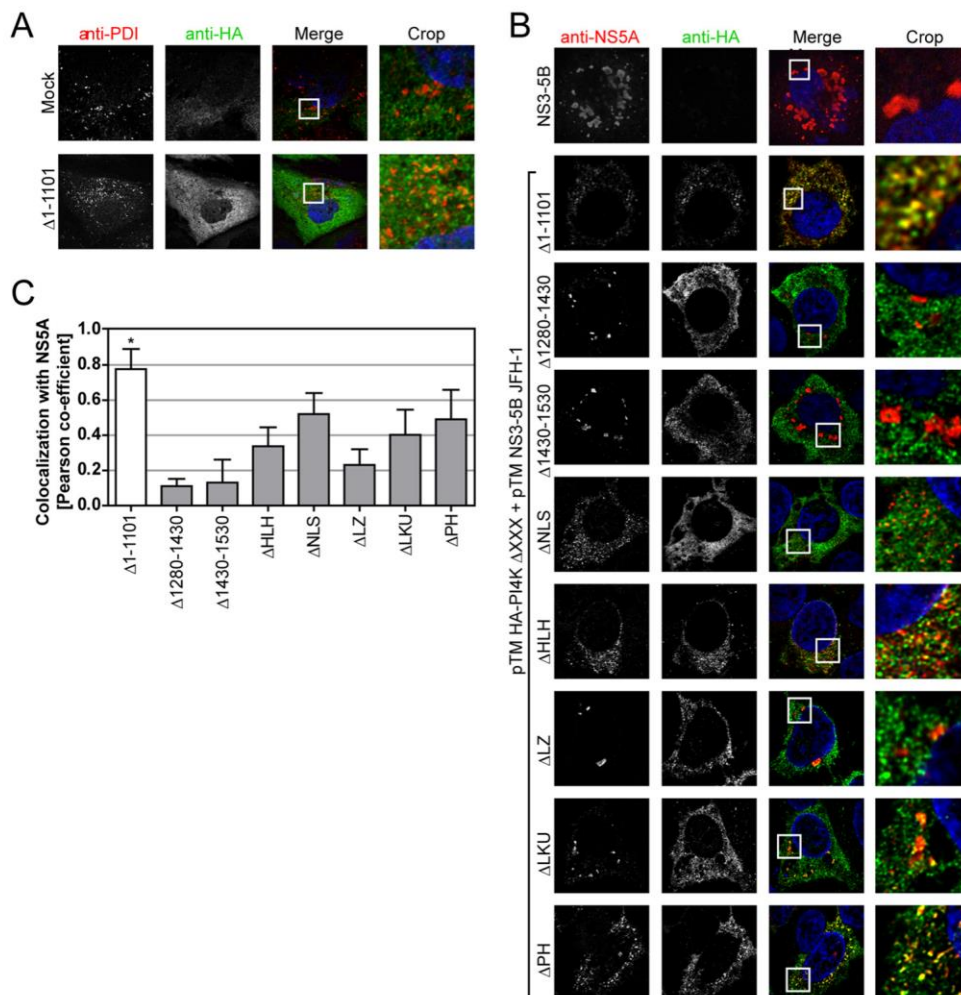




**FIG 7** Enzymatic activity of PI4KIIIα Δ1-1101 proteins harboring additional subdomain deletions. (A) Plasmids encoding HCV NS3-5B were transfected into Huh7-Lunet T7 cell lines depleted of endogenous PI4KIIIα and stably expressing the indicated variants of PI4KIIIα upon lentiviral transduction (for details, refer to the legend of Fig. 2). Twenty-four h after transfection, cells were fixed and permeabilized with digitonin. NS5A (red) or PI4P (green) was stained using monoclonal antibodies. Nuclei (blue) were stained using DAPI. (B) Fluorescence signals of PI4P from whole-cell z-stacks were quantified using ImageJ software. PI4P induction was quantified by normalizing intensities of HCV-transfected cells to mock-transfected cells of each cell line. Standard deviations are indicated. At least 30 cells were quantified for each condition. (C) Plasmids encoding the indicated HA-tagged PI4KIIIα mutants were transfected into Huh7-Lunet T7 shPI4K cells. Twenty-four h after transfection, cells were lysed and HA-tagged proteins were immunoprecipitated using HA beads. PI4KIIIα and β-actin were visualized by Western blotting using antibodies against the HA peptide and β-actin, respectively, in the cell lysate (Input) and after IP. The ca. 46-kDa band in the IP fractions corresponds to the heavy chain of the antibodies used for IP. (D) Immunoprecipitated HA-PI4KIIIα mutants were subjected to *in vitro* lipid kinase assays based on incorporation of <sup>32</sup>P into phosphatidylinositol substrates. Radioactive lipids were extracted and subjected to liquid scintillation counting. Depicted are counts per minute (cpm) obtained for the different PI4KIIIα mutants after background subtraction. Note that recovery rates were reproducibly low for mutants ΔLKU and ΔPH; therefore, low levels of residual lipid kinase activity formally cannot be excluded.

described as an armadillo repeat (ARM) domain (aa 1536 to 1723, encompassing LKU) followed by a two-lobed kinase domain (N-lobe, aa 1724 to 1900; C-lobe, aa 1901 to 2102) (Fig. 9C). Thus, the N-lobe's N terminus corresponds to the region described by Gehrmann et al. as a PH domain (13), although it lacks the two N-terminal strands and the C-terminal capping helix of the pleckstrin fold. Instead, it is flanked by two helices at its N-terminal end

and one helix and two strands at its C-terminal end. The last strand terminates in a loop that crosses to the C-lobe and provides the floor to the ATP- and wortmannin-binding pocket that is flanked by both lobes (Fig. 9D). The C-lobe harbors the catalytic residues, including D1957 (Fig. 9D), and the activation loop (Fig. 9D, represented in magenta) that is involved in PI recognition and contributes to determining phosphoinositide specificity (48). Rel-



**FIG 8** Subcellular localization of PI4KIII $\alpha$  domain deletion mutants in the presence and absence of HCV NS proteins. Huh7-Lunet shPI4K cells were transfected with plasmids encoding the indicated HA-tagged PI4KIII $\alpha$  variants (A, B, and C) and with constructs expressing HCV NS3 to NS5B (B and C). (A) Proteins were stained using monoclonal antibodies directed against PDI as the marker of the ER (red) or HA-peptide (green), respectively. (B) HCV NS5A or HA-tagged proteins were stained using monoclonal antibodies directed against NS5A (red) or HA peptide (green), respectively. (C) 3D colocalization of NS5A and PI4KIII $\alpha$  in experiments shown in panel B was assessed by recording whole-cell z-stacks and analyzed using Huygens Essential software. Depicted are mean values and standard deviations of Pearson correlation coefficients from at least 10 individual cells per condition. \*,  $P < 0.05$ .

ative to the catalytic residues and the activation loop (and the ER membrane), the 1810-1840 insertion lies on the opposite face of PI4KIII $\alpha$  next to a region that, in PI3K, is occupied by regulatory subdomains (Fig. 9C, region represented by a yellow ellipse).

In this study, we provide an updated view on the C terminus of PI4KIII $\alpha$  (summarized in Fig. 9A) that can guide further studies on the roles of different C-terminal domains in the function of PI4KIII $\alpha$  and also of related lipid kinases.

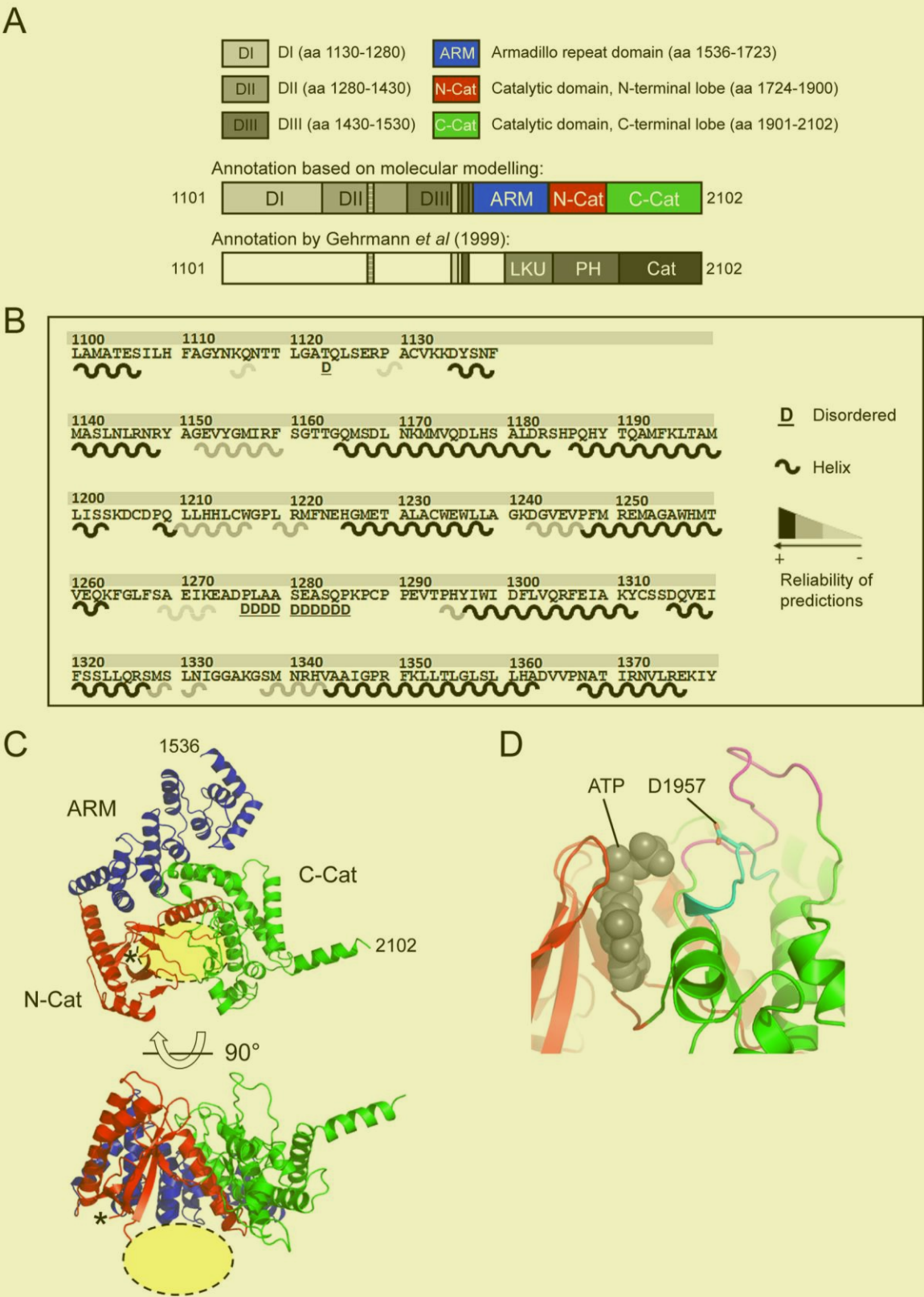
## DISCUSSION

In this study, we investigated structural requirements of PI4KIII $\alpha$  for HCV replication and enzymatic activity. Therefore, we generated a panel of PI4KIII $\alpha$  deletion mutants and characterized them regarding their enzymatic activity, their ability to be activated by expression of the HCV polyprotein, their efficiency in rescuing HCV replication, their subcellular localization in the presence and absence of HCV proteins, and their interaction with HCV NS5A. We found a very close linkage between enzymatic activity, induc-

tion of PI4P, and rescue of HCV replication, in line with our previous results and those of others, suggesting that the catalytically active center of PI4KIII $\alpha$  is essential for its role in HCV replication (10, 15). These data were largely connected to the ability of PI4KIII $\alpha$  to be recruited to the HCV replication sites. In contrast, most mutants containing gross nonoverlapping deletions still physically interacted with NS5A and showed a subcellular distribution unaltered from that of the full-length protein in the absence of HCV, arguing for several full-length sites and redundant determinants for PI4KIII $\alpha$  localization, not all of them being essential for function.

Our previous studies have shown that PI4KIII $\alpha$  physically interacted with HCV NS5A and NS5B (7, 10) and identified a region at the C terminus of NS5A domain I that is critically involved in the functional interaction of NS5A and PI4KIII $\alpha$  and is required for activation of PI4KIII $\alpha$  (10). We now aimed to identify the specific interaction region within PI4KIII $\alpha$ , but our results indicate that several redundant sites of physical interaction between





**FIG 9** Homology modeling of the C terminus of PI4KIII $\alpha$ . (A) Comparison of structural annotations of C-terminal PI4KIII $\alpha$  domains by Gehrmann *et al.* (13) and of results from homology modeling using PI3K as the template (48). All numberings refer to accession number NP\_477352.3. The region 1101-1535 cannot be modeled reliably, but secondary structure predictions suggest three helical subdomains (DI to DIII). (B) Secondary structure prediction of helical domain DI. Predicted helices are shown as waves, and gray values correspond to the probability of helix formation. Regions that tend to disorder are marked with an underlined D. (C) Homology model of PI4KIII $\alpha$  aa 1536 to 2102 colored according to the nomenclature established for the homologous PI3K. Blue, helical

PI4KIII $\alpha$  and NS5A exist, not all of which are necessarily functionally important. A previous study has shown that PI4KIII $\alpha$  interacted with NS5A through amino acids 400 to 600 (35), but it seems very unlikely that this interaction is required to support HCV replication, since deletion of this region did not compromise the ability to rescue HCV replication ( $\Delta$ 1-1101) (Table 1). Since all deletion constructs we have analyzed still coprecipitated to some extent with NS5A (and NS5B; data not shown), it is hard to assign a clearly defined region involved in functional HCV interaction. Interestingly, we found some discrepancy between sole expression of NS5A and that of the NS3-5B polyprotein. This might point to a modulation of the NS5A-PI4KIII $\alpha$  interaction by NS5A phosphorylation, since proper phosphorylation of NS5A requires expression of at least NS3-5A as a polyprotein (49, 50). Modulation of NS5A interaction by phosphorylation has been demonstrated for VAP-A (51) and might be a reasonable concept for PI4KIII $\alpha$ , since in a previous study we have shown that PI4KIII $\alpha$  directly or indirectly modulates NS5A phosphorylation (10). Although we cannot reach a final conclusion, some evidence points to an important role of the region 1102-1152 in mediating the functional interactions between NS5A and PI4KIII $\alpha$ , since mutant  $\Delta$ 1-1152 was already slightly impaired in specific recruitment to the HCV replication sites while still retaining full enzymatic activity and the ability to rescue HCV replication. Interestingly, specific recruitment was lost for many mutants lacking enzymatic activity (Table 1), suggesting that a functional interaction with HCV proteins requires not only specific sequence motifs but also a properly folded PI4KIII $\alpha$  enzyme. Still, enzymatic activity was not required for recruitment/interaction, since the inactive PI4KIII $\alpha$  point mutant D1957A interacts and colocalizes with NS5A as the wt enzyme. These findings suggest that recruitment of PI4KIII $\alpha$  by NS5A to defined compartments precedes stimulation of kinase activity. Clearly, a more subtle mutational analysis of the C-terminal half of PI4KIII $\alpha$  will be required to unravel the sites important for functional interaction with NS5A.

PI4KIII $\alpha$  has been identified as an essential host factor of HCV RNA replication by a multitude of studies (7, 11, 27–32), but the precise mechanism of action still remains to be determined. It remains difficult to dissect specific functions of PI4KIII $\alpha$  in HCV replication, since all features of PI4KIII $\alpha$  we assessed in this study were tightly linked. All mutants analyzed in this study containing detectable lipid kinase activity gave rise to increased intracellular PI4P levels and were able to rescue HCV replication. These data suggest that HCV exploits the general regulatory mechanisms governing the lipid kinase activity of PI4KIII $\alpha$ , which have not been identified yet (16). In addition, our data clearly show that PI4KIII $\alpha$ , even though it is essential for viral replication, is not a limiting factor in Huh-7 cells, since already-low endogenous amounts of PI4KIII $\alpha$  were sufficient to allow maximal viral replication efficiency which could not be further increased by ectopic overexpression.

Besides analyzing aspects regarding HCV replication, for the

first time our study provides a comprehensive deletion analysis to define functional domains of PI4KIII $\alpha$  important for enzymatic activity. Earlier studies identified an enzymatically active mutant which lacked the N-terminal 872 amino acids, while deletion of 1,189 amino acids entirely abrogated enzymatic function (corresponding to residues 931 and 1249, respectively, in our PI4KIII $\alpha$  gene), arguing for a 130-kDa enzyme as a minimal requirement for enzymatic activity (20, 52, 53), whereas an alternative splice variant of PI4KIII $\alpha$  lacking 1,249 N-terminal amino acids was shown to be inactive and did not support HCV replication (15). Our deletion analysis now revealed that the N-terminal 1,152 amino acids are dispensable for enzymatic activity, whereas further deletion of 50 amino acids resulted in an inactive protein ( $\Delta$ 1-1202). These results more closely define the PI4KIII $\alpha$  enzymatic core to a size of ca. 105 to 108 kDa, which is surprisingly similar to that of related enzymes like PI3K (48). Interestingly, we did not find evidence for a function of the N-terminal half of PI4KIII $\alpha$  in any of our assay systems regarding subcellular localization, enzymatic activity, and stimulation by HCV. A previous study identified the very N-terminal 58 amino acids involved in dynamic shuttling of PI4KIII $\alpha$  to the plasma membrane (22), a sequence that is missed in most GenBank entries. However, since we used a construct of length identical to the one used by Nakatsu et al., plasma membrane shuttling should be retained with our full-length enzyme but lost for  $\Delta$ 1-1101, thereby indicating that this mechanism is not essential for enzymatic activity or HCV replication.

Surprisingly, all of our deletion mutants showed a subcellular distribution comparable to that of the full-length enzyme in the absence of HCV proteins. The majority of signal was broadly localized throughout the cytoplasm, in line with published data (18, 22). Even the very N terminus ( $\Delta$ 802-2103) as well as the very C terminus ( $\Delta$ 1-1794) showed no gross change in subcellular localization compared to the full-length protein. Therefore, we were not able to assign a specific sequence within PI4KIII $\alpha$  responsible for subcellular localization. In previous reports, we and others found a more distinct localization of the endogenous kinase in punctate structures throughout the cytoplasm, consistent with an ER association (7, 16, 21). In brain cells, PI4KIII $\alpha$  was prominently detected in the nucleus and nucleolus (21). Therefore, the subcellular localization might depend on cell type and expression level of PI4KIII $\alpha$ . Although the majority of PI4KIII $\alpha$  has been found associated with membranes (18), the determinants of membrane association have not been found yet for the mammalian enzymes (16). In light of the very recent observation of a dynamic recruitment of PI4KIII $\alpha$  to the plasma membrane mediated by a protein complex containing EFR3 (22), it is tempting to speculate that PI4KIII $\alpha$  is recruited to the ER as well by auxiliary factors, such as VAP-B (22), and not by an intrinsic function of the kinase. In this scenario, overexpression like that in our experiments might overload the membrane recruitment complex, resulting in a broad pleiotropic localization. In line with this idea,

armadillo repeat domain (ARM); red and green, kinase domain, N-lobe (N-Cat) and C-lobe (C-Cat), respectively. The 1810-1840 region of PI4KIII $\alpha$  (not modeled) would be inserted at the location marked by an asterisk. A location below the two lobes that is often occupied by regulatory and recruitment domains N terminal to the homologous ARM domain in PI3K is highlighted as a yellow ellipse. A view rotated by 90° around the horizontal axis is given at the bottom. Note that aa 2103 in our PI4KIII $\alpha$  variant is missing from all other GenBank entries; therefore, it was not included in the modeling. (D) Zoom on the interface between the N-lobe and C-lobe. The ATP pocket that is also the target of wortmannin is highlighted with a modeled ATP in gray spheres. The catalytic loop and activation loop in the C-lobe are shown in cyan and magenta, respectively. The catalytic residue D1957 in the catalytic loop is displayed as sticks.



the majority of PI4KIII $\alpha$  was recruited to NS5A and vanished from the cytoplasm upon co-overexpression of HCV NS3 to NS5B, which forms an abundant membrane-associated complex interacting with the kinase and probably competing with the cellular recruitment complex. However, further experimentation is required to prove this hypothesis, including the assessment of the membrane-associated fractions of PI4KIII $\alpha$  wt and deletion mutants and a proteomic analysis of potential PI4KIII $\alpha$  interaction partners in the ER.

Since our N-terminal deletion mutant,  $\Delta$ 1-1101, was comparable to the full-length enzyme in all of our assay systems, we focused our further deletion mapping on domains and motifs predicted to be located in the C-terminal half of PI4KIII $\alpha$  (14). Although our experimental models were set out to differentiate between different functions, again we found a tight linkage of almost all phenotypes, except subcellular localization in the presence of HCV proteins. With the exception of  $\Delta$ NLS, which was only moderately impaired in enzymatic activity, all other deletion mutants were completely devoid of lipid kinase activity *in vitro* and activation by HCV. While this result might be expected for large deletions like LKU, it seems very unlikely that a short motif like HLH or LZ is essential for the enzymatic function. HLH and LZ in particular are supposed to have a role in the recognition of DNA sequences, probably for the subfraction of PI4KIII $\alpha$  located in the nucleus (54). The loss of function associated with all N-terminal mutants beyond aa 1152 and internal deletions, as well as the impaired function of the  $\Delta$ NLS mutant, might point to the fact that these predicted motifs are located in regions relevant for the overall structure of an active enzyme. Therefore, our deletion analysis does not rule out the existence of these predicted motifs, and more subtle point mutations will be required to distinctively address their roles in enzymatic function. Still, the results of our deletion analysis are in favor of our structural model, which is based on homology to the known structure of PI3K (48) and predictions (Fig. 9). This model suggests that the region previously designated the PH domain builds the N-terminal lobe of the catalytic core, whereas the LKU domain forms an armadillo repeat. Armadillo domains mediate protein-protein interactions and are scaffold domains. Thus, in PI3K they buttress the membrane-interacting C2 domain and noncatalytic subunit. Furthermore, our data suggest that the region between aa 1152 and 1530 is, in large part, essential for enzymatic activity and is organized into three distinct alpha-helical subdomains (Fig. 9). Disruption of these secondary structures by deletion as well as N-terminal truncations beyond aa 1152 and even deletions of smaller parts in DIII resulted in a complete loss of enzymatic activity. Only the residual activity found with  $\Delta$ NLS argues for some structural flexibility in DII. It is tempting to speculate that domains DI to DIII are involved in the regulation of enzymatic activity. Allosteric regulation of kinase activity has been described before for the PI3K regulatory subunit p85 that binds and stabilizes its catalytic subunit, p110 (55). Upon stimulation, e.g., by phosphorylation or by binding of phosphoproteins to SH2 domains of p85, enzymatic activity of the p110 subunit is increased (56). Binding of interaction partners like the ADP-ribosylation factor also stimulates functions of the related PI4KIII $\beta$  isoform (57). Activation by phosphorylation also has been demonstrated for PI4KIII $\beta$  that is phosphorylated by protein kinase D at a highly conserved motif, resulting in stimulation of lipid kinase activity (58). Interestingly, the phosphorylated residue lies in an insertion between the PI4KIII $\beta$  ARM and

N-CAT that, in the PI4KIII $\alpha$  model, is right next to the 1810-1840 insertion (Fig. 9C). In PI3K, this region opposite the membrane and below the floor of the catalytic site (Fig. 9C, yellow ellipse) is occupied by regulatory domains. For instance, the Ras-binding domain (RBD) of PI3K $\gamma$  is located in this region. Binding of Ras across RBD and C-CAT induces a conformational change involving both membrane-interacting and catalytic regions of PI3K $\gamma$ , suggesting an allosteric mechanism in the kinase activation (59). PI4KIII $\alpha$  may be similarly recruited to membranes and activated, although this assumption requires further experimental validation. With respect to specific recruitment of PI4KIII $\alpha$  by HCV rather than recruitment of PI4KIII $\beta$ , it is interesting that PI4KIII $\beta$  does not have a region N terminal to its ARM domain similar to DI-DIII of PI4KIII $\alpha$ .

Taking our results together, with this study we characterized PI4KIII $\alpha$  in depth as an essential HCV host factor and defined requirements for HCV replication and enzymatic activity, and we provided insights into the structural complexity that mediates kinase function. We provide a structural model of the catalytic center of the enzyme and lay the ground for more detailed reverse-genetic analyses to define more precisely the functions of individual PI4KIII $\alpha$  subdomains.

## ACKNOWLEDGMENTS

We especially thank Ulrike Herian for excellent technical assistance. We are grateful to Charles M. Rice for antibody 9E10, Takaji Wakita for the JFH-1 isolate, and Mark Harris for the polyclonal sheep serum recognizing NS5A.

This project was funded by grants from the Deutsche Forschungsgemeinschaft (LO 1556/1-2 and TRR77, TPA1) to V.L. and by ANRS (France Recherche Nord & sud SIDA–hiv Hépatites: FRENSh) through a grant to S.B. and a predoctoral fellowship to C.T.

## REFERENCES

- Moradpour D, Penin F, Rice CM. 2007. Replication of hepatitis C virus. *Nat. Rev. Microbiol.* 5:453–463. <http://dx.doi.org/10.1038/nrmicro1645>.
- Bartenschlager R, Cosset FL, Lohmann V. 2010. Hepatitis C virus replication cycle. *J. Hepatol.* 53:583–585. <http://dx.doi.org/10.1016/j.jhep.2010.04.015>.
- Lohmann V. 2013. Hepatitis C virus RNA replication. *Curr. Top. Microbiol. Immunol.* 369:167–198. [http://dx.doi.org/10.1007/978-3-642-27340-7\\_7](http://dx.doi.org/10.1007/978-3-642-27340-7_7).
- Egger D, Wolk B, Gosert R, Bianchi L, Blum HE, Moradpour D, Bienz K. 2002. Expression of hepatitis C virus proteins induces distinct membrane alterations including a candidate viral replication complex. *J. Virol.* 76:5974–5984. <http://dx.doi.org/10.1128/JVI.76.12.5974-5984.2002>.
- Gosert R, Egger D, Lohmann V, Bartenschlager R, Blum HE, Bienz K, Moradpour D. 2003. Identification of the hepatitis C virus RNA replication complex in Huh-7 cells harboring subgenomic replicons. *J. Virol.* 77:5487–5492. <http://dx.doi.org/10.1128/JVI.77.9.5487-5492.2003>.
- Ferraris P, Blanchard E, Roingeard P. 2010. Ultrastructural and biochemical analyses of hepatitis C virus-associated host cell membranes. *J. Gen. Virol.* 91:2230–2237. <http://dx.doi.org/10.1099/vir.0.022186-0>.
- Reiss S, Rebhan I, Backes P, Romero-Brey I, Erfle H, Matula P, Kaderali L, Poenisch M, Blankenburg H, Hiet MS, Longerich T, Diehl S, Ramirez F, Balla T, Rohr K, Kaul A, Buhler S, Pepperkok R, Lengauer T, Albrecht M, Eils R, Schirmacher P, Lohmann V, Bartenschlager R. 2011. Recruitment and activation of a lipid kinase by hepatitis C virus NS5A is essential for integrity of the membranous replication compartment. *Cell Host Microbe* 9:32–45. <http://dx.doi.org/10.1016/j.chom.2010.12.002>.
- Romero-Brey I, Merz A, Chiramel A, Lee JY, Chlanda P, Haselman U, Santarella-Mellwig R, Habermann A, Hoppe S, Kallis S, Walther P, Antony C, Krijnse-Locker J, Bartenschlager R. 2012. Three-dimensional architecture and biogenesis of membrane structures associated with hepatitis C virus replication. *PLoS Pathog.* 8:e1003056. <http://dx.doi.org/10.1371/journal.ppat.1003056>.

9. Paul D, Hoppe S, Saher G, Krijnse-Locker J, Bartenschlager R. 2013. Morphological and biochemical characterization of the membranous hepatitis C virus replication compartment. *J. Virol.* 87:10612–10627. <http://dx.doi.org/10.1128/JVI.01370-13>.
10. Reiss S, Harak C, Romero-Brey I, Radujkovic D, Klein R, Ruggieri A, Rebhan I, Bartenschlager R, Lohmann V. 2013. The lipid kinase phosphatidylinositol-4 kinase III alpha regulates the phosphorylation status of hepatitis C virus NS5A. *PLoS Pathog.* 9:e1003359. <http://dx.doi.org/10.1371/journal.ppat.1003359>.
11. Berger KL, Cooper JD, Heaton NS, Yoon R, Oakland TE, Jordan TX, Mateu G, Grakoui A, Randall G. 2009. Roles for endocytic trafficking and phosphatidylinositol 4-kinase III alpha in hepatitis C virus replication. *Proc. Natl. Acad. Sci. U. S. A.* 106:7577–7582. <http://dx.doi.org/10.1073/pnas.0902693106>.
12. D'Angelo G, Vicinanza M, Di Campli A, De Matteis MA. 2008. The multiple roles of PtdIns(4)P—not just the precursor of PtdIns(4,5)P<sub>2</sub>. *J. Cell Sci.* 121:1955–1963. <http://dx.doi.org/10.1242/jcs.023630>.
13. Gehrman T, Gulkan H, Suer S, Herberg FW, Balla A, Vereb G, Mayr GW, Heilmeyer LM, Jr. 1999. Functional expression and characterization of a new human phosphatidylinositol 4-kinase PI4K230. *Biochim. Biophys. Acta* 1437:341–356. [http://dx.doi.org/10.1016/S1388-1981\(99\)00029-3](http://dx.doi.org/10.1016/S1388-1981(99)00029-3).
14. Gehrman T, Heilmeyer LM, Jr. 1998. Phosphatidylinositol 4-kinases. *Eur. J. Biochem.* 253:357–370. <http://dx.doi.org/10.1046/j.1432-1327.1998.2530357.x>.
15. Szentpetery Z, Szakacs G, Bojjireddy N, Tai AW, Balla T. 2011. Genetic and functional studies of phosphatidylinositol 4-kinase type IIIalpha. *Biochim. Biophys. Acta* 1811:476–483. <http://dx.doi.org/10.1016/j.bbalip.2011.04.013>.
16. Balla A, Balla T. 2006. Phosphatidylinositol 4-kinases: old enzymes with emerging functions. *Trends Cell Biol.* 16:351–361. <http://dx.doi.org/10.1016/j.tcb.2006.05.003>.
17. Audhya A, Emr SD. 2002. Stt4 PI 4-kinase localizes to the plasma membrane and functions in the Pkc1-mediated MAP kinase cascade. *Dev. Cell* 2:593–605. [http://dx.doi.org/10.1016/S1534-5807\(02\)00168-5](http://dx.doi.org/10.1016/S1534-5807(02)00168-5).
18. Wong K, Meyers D, Cantley LC. 1997. Subcellular locations of phosphatidylinositol 4-kinase isoforms. *J. Biol. Chem.* 272:13236–13241. <http://dx.doi.org/10.1074/jbc.272.20.13236>.
19. Nakagawa T, Goto K, Kondo H. 1996. Cloning, expression, and localization of 230-kDa phosphatidylinositol 4-kinase. *J. Biol. Chem.* 271:12088–12094. <http://dx.doi.org/10.1074/jbc.271.20.12088>.
20. Heilmeyer LM, Jr, Vereb G, Jr, Vereb G, Kakuk A, Szivak I. 2003. Mammalian phosphatidylinositol 4-kinases. *IUBMB Life* 55:59–65. <http://dx.doi.org/10.1002/tbmb.718540873>.
21. Kakuk A, Friedlander E, Vereb G, Jr, Kasa A, Balla A, Balla T, Heilmeyer LM, Jr, Gergely P, Vereb G. 2006. Nucleolar localization of phosphatidylinositol 4-kinase PI4K230 in various mammalian cells. *Cytometry A* 69:1174–1183. <http://dx.doi.org/10.1002/cyto.a.20347>.
22. Nakatsu F, Baskin JM, Chung J, Tanner LB, Shui G, Lee SY, Pirruccello M, Hao M, Ingolia NT, Wenk MR, De Camilli P. 2012. PtdIns4P synthesis by PI4KIIIalpha at the plasma membrane and its impact on plasma membrane identity. *J. Cell Biol.* 199:1003–1016. <http://dx.doi.org/10.1083/jcb.201206095>.
23. Balla A, Tuymetova G, Tsiomenko A, Varnai P, Balla T. 2005. A plasma membrane pool of phosphatidylinositol 4-phosphate is generated by phosphatidylinositol 4-kinase type-III alpha: studies with the PH domains of the oxysterol binding protein and FAPP1. *Mol. Biol. Cell* 16:1282–1295. <http://dx.doi.org/10.1091/mbc.E04-07-0578>.
24. Bianco A, Reghellin V, Donnici L, Fenu S, Alvarez R, Baruffa C, Peri F, Pagani M, Abbrignani S, Neddermann P, De Francesco R. 2012. Metabolism of phosphatidylinositol 4-kinase IIIalpha-dependent PI4P is subverted by HCV and is targeted by a 4-anilino quinazoline with antiviral activity. *PLoS Pathog.* 8:e1002576. <http://dx.doi.org/10.1371/journal.ppat.1002576>.
25. Wong K, Cantley LC. 1994. Cloning and characterization of a human phosphatidylinositol 4-kinase. *J. Biol. Chem.* 269:28878–28884. <http://dx.doi.org/10.1074/jbc.271.20.12088>.
26. Altan-Bonnet N, Balla T. 2012. Phosphatidylinositol 4-kinases: hostages harnessed to build panviral replication platforms. *Trends Biochem. Sci.* 37:293–302. <http://dx.doi.org/10.1016/j.tibs.2012.03.004>.
27. Hsu NY, Ilynska O, Belov G, Santiana M, Chen YH, Takvorian PM, Pau C, van der Schaar H, Kaushik-Basu N, Balla T, Cameron CE, Ehrenfeld E, van Kuppeveld FJ, Altan-Bonnet N. 2010. Viral reorganization of the secretory pathway generates distinct organelles for RNA replication. *Cell* 141:799–811. <http://dx.doi.org/10.1016/j.cell.2010.03.050>.
28. Tai AW, Benita Y, Peng LF, Kim SS, Sakamoto N, Xavier RJ, Chung RT. 2009. A functional genomic screen identifies cellular cofactors of hepatitis C virus replication. *Cell Host Microbe* 5:298–307. <http://dx.doi.org/10.1016/j.chom.2009.02.001>.
29. Li Q, Brass AL, Ng A, Hu Z, Xavier RJ, Liang TJ, Elledge SJ. 2009. A genome-wide genetic screen for host factors required for hepatitis C virus propagation. *Proc. Natl. Acad. Sci. U. S. A.* 106:16410–16415. <http://dx.doi.org/10.1073/pnas.0907439106>.
30. Trotard M, Lepere-Douard C, Regeard M, Piquet-Pellorce C, Lavillette D, Cosset FL, Gripon P, Le Seyec J. 2009. Kinases required in hepatitis C virus entry and replication highlighted by small interference RNA screening. *FASEB J.* 23:3780–3789. <http://dx.doi.org/10.1096/fj.09-131920>.
31. Vaillancourt FH, Pilote L, Cartier M, Lippens J, Liuzzi M, Bethell RC, Cordingley MG, Kukolj G. 2009. Identification of a lipid kinase as a host factor involved in hepatitis C virus RNA replication. *Virology* 387:5–10. <http://dx.doi.org/10.1016/j.virol.2009.02.039>.
32. Borawski J, Troke P, Puyang X, Gibaja V, Zhao S, Mickanin C, Leighton-Davies J, Wilson CJ, Myer V, Cornellataracido I, Baryza J, Tallarico J, Joberty G, Bantscheff M, Schirle M, Bouwmeester T, Mathy JE, Lin K, Compton T, Labow M, Wiedmann B, Gaither LA. 2009. Class III phosphatidylinositol 4-kinase alpha and beta are novel host factor regulators of hepatitis C virus replication. *J. Virol.* 83:10058–10074. <http://dx.doi.org/10.1128/JVI.02418-08>.
33. Berger KL, Kelly SM, Jordan TX, Tartell MA, Randall G. 2011. Hepatitis C virus stimulates the phosphatidylinositol 4-kinase III alpha-dependent phosphatidylinositol 4-phosphate production that is essential for its replication. *J. Virol.* 85:8870–8883. <http://dx.doi.org/10.1128/JVI.00059-11>.
34. Wang H, Perry JW, Laurant AS, Neddermann P, De Francesco R, Tai AW. 2014. Oxysterol-binding protein is a phosphatidylinositol 4-kinase effector required for HCV replication membrane integrity and cholesterol trafficking. *Gastroenterology* 146:1373–1385. <http://dx.doi.org/10.1053/j.gastro.2014.02.002>.
35. Lim YS, Hwang SB. 2011. Hepatitis C virus NS5A protein interacts with phosphatidylinositol 4-kinase type III alpha and regulates viral propagation. *J. Biol. Chem.* 286:11290–11298. <http://dx.doi.org/10.1074/jbc.M110.194472>.
36. Backes P, Quinkert D, Reiss S, Binder M, Zayas M, Rescher U, Gerke V, Bartenschlager R, Lohmann V. 2010. Role of annexin A2 in the production of infectious hepatitis C virus particles. *J. Virol.* 84:5775–5789. <http://dx.doi.org/10.1128/JVI.02343-09>.
37. Simister P, Schmitt M, Geitmann M, Wicht O, Danielson UH, Klein R, Bressanelli S, Lohmann V. 2009. Structural and functional analysis of hepatitis C virus strain JFH1 polymerase. *J. Virol.* 83:11926–11939. <http://dx.doi.org/10.1128/JVI.01008-09>.
38. Binder M, Quinkert D, Bochkarova O, Klein R, Kezmic N, Bartenschlager R, Lohmann V. 2007. Identification of determinants involved in initiation of hepatitis C virus RNA synthesis by using intergenotypic replicase chimeras. *J. Virol.* 81:5270–5283. <http://dx.doi.org/10.1128/JVI.00032-07>.
39. Altshul SJ, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389–3402. <http://dx.doi.org/10.1093/nar/25.17.3389>.
40. Papadopoulos JS, Agarwala R. 2007. COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics* 23:1073–1079. <http://dx.doi.org/10.1093/bioinformatics/btm076>.
41. Roy A, Kucukural A, Zhang Y. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* 5:725–738. <http://dx.doi.org/10.1038/nprot.2010.5>.
42. Miller S, Tavshanjan B, Oleksy A, Perisic O, Houseman BT, Shokat KM, Williams RL. 2010. Shaping development of autophagy inhibitors with the structure of the lipid kinase Vps34. *Science* 327:1638–1642. <http://dx.doi.org/10.1126/science.1184429>.
43. Biegert A, Mayer C, Remmert M, Soding J, Lupas AN. 2006. The MPI Bioinformatics Toolkit for protein sequence analysis. *Nucleic Acids Res.* 34:W335–W339. <http://dx.doi.org/10.1093/nar/gkl217>.
44. Wang R, Brattain MG. 2007. The maximal size of protein to diffuse through the nuclear pore is larger than 60kDa. *FEBS Lett.* 581:3164–3170. <http://dx.doi.org/10.1016/j.febslet.2007.05.082>.
45. Hildebrand A, Remmert M, Biegert A, Soding J. 2009. Fast and accurate



- automatic structure prediction with HHpred. *Proteins* 77(Suppl 9):128–132. <http://dx.doi.org/10.1002/prot.22499>.
46. Brown JR, Auger KR. 2011. Phylogenomics of phosphoinositide lipid kinases: perspectives on the evolution of second messenger signaling and drug discovery. *BMC Evol. Biol.* 11:4. <http://dx.doi.org/10.1186/1471-2148-11-4>.
  47. Gehrmann T, Vereb G, Schmidt M, Klix D, Meyer HE, Varsanyi M, Heilmeyer LM, Jr. 1996. Identification of a 200 kDa polypeptide as type 3 phosphatidylinositol 4-kinase from bovine brain by partial protein and cDNA sequencing. *Biochim. Biophys. Acta* 1311:53–63.
  48. Walker EH, Perisic O, Ried C, Stephens L, Williams RL. 1999. Structural insights into phosphoinositide 3-kinase catalysis and signalling. *Nature* 402:313–320. <http://dx.doi.org/10.1038/46319>.
  49. Koch JO, Bartenschlager R. 1999. Modulation of hepatitis C virus NS5A hyperphosphorylation by nonstructural proteins NS3, NS4A, and NS4B. *J. Virol.* 73:7138–7146.
  50. Neddermann P, Clementi A, De Francesco R. 1999. Hyperphosphorylation of the hepatitis C virus NS5A protein requires an active NS3 protease, NS4A, NS4B, and NS5A encoded on the same polyprotein. *J. Virol.* 73:9984–9991.
  51. Evans MJ, Rice CM, Goff SP. 2004. Phosphorylation of hepatitis C virus nonstructural protein 5A modulates its protein interactions and viral RNA replication. *Proc. Natl. Acad. Sci. U. S. A.* 101:13038–13043. <http://dx.doi.org/10.1073/pnas.0405152101>.
  52. Vaillancourt FH, Brault M, Pilote L, Uyttersprot N, Gaillard ET, Stoltz JH, Knight BL, Pantages L, McFarland M, Breitfelder S, Chiu TT, Mahrouche L, Faucher AM, Cartier M, Cordingley MG, Bethell RC, Jiang H, White PW, Kukolj G. 2012. Evaluation of phosphatidylinositol-4-kinase IIIalpha as a hepatitis C virus drug target. *J. Virol.* 86:11595–11607. <http://dx.doi.org/10.1128/JVI.01320-12>.
  53. Knight ZA, Gonzalez B, Feldman ME, Zunder ER, Goldenberg DD, Williams O, Loewith R, Stokoe D, Balla A, Toth B, Balla T, Weiss WA, Williams RL, Shokat KM. 2006. A pharmacological map of the PI3-K family defines a role for p110alpha in insulin signaling. *Cell* 125:733–747. <http://dx.doi.org/10.1016/j.cell.2006.03.035>.
  54. Kakuk A, Friedlander E, Vereb G, Jr, Lisboa D, Bagossi P, Toth G, Gergely P, Vereb G. 2008. Nuclear and nucleolar localization signals and their targeting function in phosphatidylinositol 4-kinase PI4K230. *Exp. Cell Res.* 314:2376–2388. <http://dx.doi.org/10.1016/j.yexcr.2008.05.006>.
  55. Yu J, Zhang Y, McIlroy J, Rordorf-Nikolic T, Orr GA, Backer JM. 1998. Regulation of the p85/p110 phosphatidylinositol 3'-kinase: stabilization and inhibition of the p110alpha catalytic subunit by the p85 regulatory subunit. *Mol. Cell. Biol.* 18:1379–1387.
  56. Shoelson SE, Sivaraja M, Williams KP, Hu P, Schlessinger J, Weiss MA. 1993. Specific phosphopeptide binding regulates a conformational change in the PI 3-kinase SH2 domain associated with enzyme activation. *EMBO J.* 12:795–802.
  57. Godi A, Pertile P, Meyers R, Marra P, Di TG, Iurisci C, Luini A, Corda D, De Matteis MA. 1999. ARF mediates recruitment of PtdIns-4-OH kinase-beta and stimulates synthesis of PtdIns(4,5)P2 on the Golgi complex. *Nat. Cell Biol.* 1:280–287. <http://dx.doi.org/10.1038/12993>.
  58. Hausser A, Storz P, Martens S, Link G, Toker A, Pfizenmaier K. 2005. Protein kinase D regulates vesicular transport by phosphorylating and activating phosphatidylinositol-4 kinase IIIbeta at the Golgi complex. *Nat. Cell Biol.* 7:880–886. <http://dx.doi.org/10.1038/ncb1289>.
  59. Pacold ME, Suire S, Perisic O, Lara-Gonzalez S, Davis CT, Walker EH, Hawkins PT, Stephens L, Eccleston JF, Williams RL. 2000. Crystal structure and functional analysis of Ras binding to its effector phosphoinositide 3-kinase gamma. *Cell* 103:931–943. [http://dx.doi.org/10.1016/S0092-8674\(00\)00196-3](http://dx.doi.org/10.1016/S0092-8674(00)00196-3).

## 2) ARM-2102, étude détaillée et comparaison avec les homologues de structure connue

*La numérotation de la PI4KA humaine est utilisée tout au long de ce chapitre pour des facilités de compréhension.*

### **i) Domaine ARM**

Ce domaine est strictement conservé chez les PI4KA de vertébrés d'une part, et des insectes d'autre part. A contrario, le nombre d'hélices composant ce domaine ainsi que la séquence primaire sont variables entre les vertébrés et les autres eucaryotes. Le domaine ARM des PI4KAs apparaît plus long que chez les PI3Ks. Or, les domaines ARM sont connus pour être impliqués dans des interactions protéines-protéines. Ainsi, un domaine ARM plus étendu peut traduire un besoin en surface d'interaction plus élevé que chez les PI3Ks ou encore une nécessité de plasticité importante concernant les partenaires d'interaction.

Proche de l'interface ARM-CAT se situe une boucle comprise entre les résidus 1697-1708 qui est conservée chez les PI4KAs (en cyan figure 22). Cette boucle possède un motif central -DEE- faiblement variable, mais toujours négativement chargé ainsi qu'une arginine conservée chez les eucaryotes supérieurs. Cette boucle n'a pas d'équivalent chez PI4KB, cependant les PI3Ks de classe I masquent la boucle correspondante par leur module C2 interagissant spécifiquement avec les phospholipides. Or, les boucles accessibles au solvant d'un tel module interagissent avec les membranes et sont composées de résidus chargés de type aspartate, glutamate, arginine et lysine tout comme la boucle observée chez les PI4KAs. Cette boucle pourrait donc être un motif d'interaction avec les membranes chez PI4KA.

En amont de cette boucle se situent deux hélices bien conservées entre PI4KA et STT4p (résidus 1659-1691, notés H1 et H2 figure 22). L'hélice 1 (H1) de ce faisceau fait le lien entre le domaine ARM et le lobe N-terminal du domaine catalytique. L'hélice 2 (H2) quant à elle, et plus particulièrement les résidus aromatiques la composant, est accessible au solvant. Or cette hélice chez la PI4KB humaine interagit entre autre avec la protéine régulatrice Rab 11<sup>57</sup>. Le domaine ARM est globalement très peu conservé en séquence entre PI4KA, STT4p, PI4KB et Pik1p. Cependant, deux résidus de cette double hélice le sont strictement :

- La **sérine 1680** (notée (\*)) sur la figure 22), positionnée dans la boucle reliant les hélices 1 et 2, et formant des liaisons hydrogènes avec chacune d'elle. Elle se situe à proximité de l'insertion 1810-1840 étudiée plus en détail par la suite.
- La **glutamine 1686**, enfouie et située sur l'hélice 1. Elle est à proximité du lobe N-terminal du domaine catalytique. Chez PI4KB, elle forme une liaison hydrogène avec une asparagine elle-même impliquée dans une liaison hydrogène avec un tryptophane situé sur une boucle du lobe catalytique N-terminal (résidus en jaune figure 22).

En ce qui concerne PI4KA, l'asparagine est remplacée par la glutamine 1681, et le tryptophane par la tyrosine 1775.

Tous ces éléments nous mènent à penser que les hélices H1 et H2 forment un site d'interaction avec des protéines et/ou des membranes. Cette interaction interviendrait dans la régulation de la kinase en provoquant des réarrangements du domaine catalytique de PI4KA par allostérie. Cette hypothèse est confortée par l'hélice en aval de ce faisceau qui est encadrée de boucles riches en prolines (orange figure 22) conservées chez PI4KA. Il s'avère que l'hélice correspondante chez PI4KB interagit aussi directement avec Rab11. De plus, la rigidification de cette hélice permettrait, lors de l'interaction avec un partenaire, un effet allostérique efficace. Une autre hypothèse quant à l'existence de ces prolines pourrait être une interaction avec un module de régulation possédant un domaine SH3, domaine couramment rencontrés dans les sous-unités régulatrices des PI3Ks.

## ii) Domaine catalytique

Le doublet d'hélices conservées H1-H2 se situe à proximité **d'une insertion (résidus 1810-1840)** située dans le lobe N-terminal du domaine catalytique. Cette insertion est inexistante chez les PI3Ks de classe I mais remplacée par une boucle plus petite permettant l'interaction avec leur module de régulation N-terminal de type ABD (Adaptator Binding Domain). Ces domaines permettent l'interaction avec les GTPases de la famille Ras à l'origine de leur activation<sup>32</sup>. Chez PI4KA, cette boucle est globalement acide (PI=4.4 pour la PI4KA humaine) et est constituée de petits résidus acides, d'arginines et d'un grand nombre de sérines conservées chez les vertébrés. Seuls les résidus 1810 à 1831 sont conservés, le reste de la séquence est variable et ce même entre primates. Chez les autres eucaryotes, cette insertion peut être plus longue mais possède toujours au moins une sérine. La boucle correspondant chez PI4KB, beaucoup plus longue, possède une sérine et une thréonine qui peuvent être phosphorylées mais pour lesquelles il n'a pas encore été déterminé de rôle<sup>104</sup>. Il est ainsi possible que cette boucle possède un rôle régulateur de l'activité de PI4KA et ce par phosphorylation et/ou par recrutement de partenaires.

Sur la base d'alignements de séquences et de modélisation, il est difficile de tirer des informations quant au domaine catalytique de PI4KA en dehors de cette insertion. En effet, la conservation est trop élevée entre PI4/3K (27% d'identité de séquences, alignement de séquence disponible en Annexe n°1). Néanmoins, la comparaison des différentes structures cristallographiques de PI3/4Ks révèle une partie du mécanisme d'action de ces kinases.





Comme attendu, la **boucle catalytique** est strictement conservée chez les PI3/4Ks avec un motif catalytique central 1957-DRHxxN-1962. L'histidine catalytique semble adopter différentes orientations corrélées à des positionnements différents de l'hélice C-terminale (voir figure 23). Ainsi, les PI3/4Ks en conformation « ouverte » présentent une histidine catalytique dirigée vers l'intérieur de la poche catalytique, formant une liaison hydrogène avec l'asparagine 1962 de la boucle catalytique. Pour les PI3Ks en conformation « fermée », l'asparagine 1962 ne peut plus former de liaison hydrogène avec l'histidine catalytique qui est dirigée vers l'hélice C-terminale de la kinase, à l'opposé de la poche catalytique. Par conséquent, la forme ouverte serait potentiellement la forme active de la kinase. En effet, dans le mécanisme proposé, l'histidine catalytique va dans un premier temps réaliser une attaque nucléophile sur le phosphate  $\gamma$  de l'ATP permettant ainsi le transfert de ce phosphate sur l'inositol du PtdIns. L'histidine catalytique doit donc être positionnée en direction de l'ATP en début de réaction catalytique.

De plus, l'asparagine 1962 semblerait intervenir dans la réaction catalytique même. En effet, après étude de la seule structure de PI3K possédant un ATP (PI3K $\gamma$ , pdb 1E8X, présentée en introduction.II.B.4), nous observons que l'asparagine 1962 interagit avec le groupement phosphate  $\gamma$  du cofacteur. Ainsi, au vu de son interaction avec l'ATP et de ses interactions possibles avec l'histidine catalytique, l'asparagine 1962 pourrait jouer un rôle fondamental dans le transfert du phosphate  $\gamma$  de l'ATP vers l'inositol. Cependant, il a été proposé que l'asparagine 1962 soit aussi à l'origine de la chélation d'un des ions divalents nécessaire à la réaction catalytique<sup>58</sup>. Ce résidu ne pourrait donc pas remplir toutes ces fonctions tout en assurant la chélation du magnésium. Or, dans la structure de PI3K $\gamma$  avec ATP (pdb 1E8X), l'aspartate 1975 est susceptible d'intervenir dans le maintien des deux magnésiums, et l'aspartate 1854, strictement conservée entre PI3/4K, permet la chélation du magnésium.

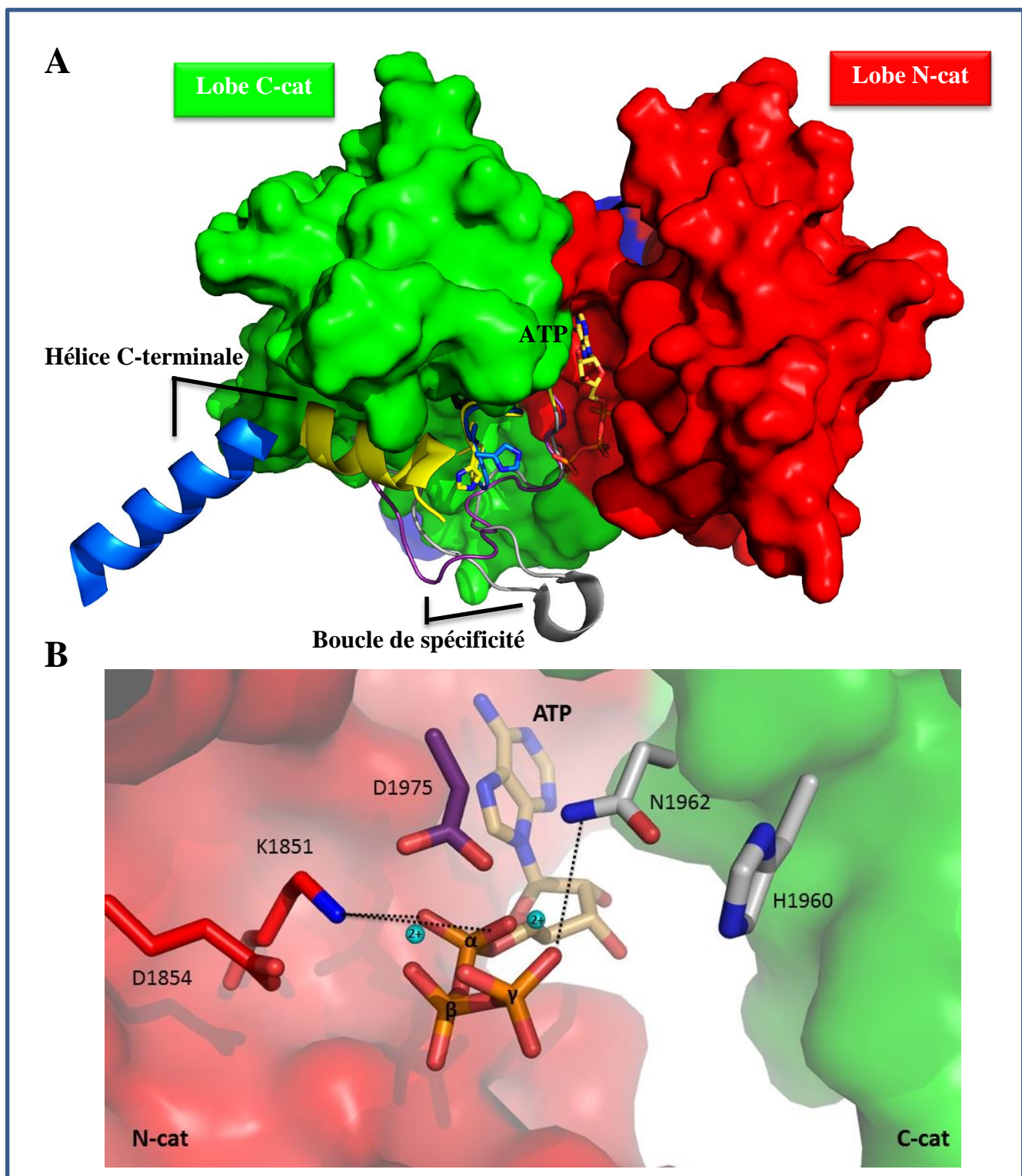
L'**hélice C-terminale**, à l'origine des configurations ouverte et fermée, est présente chez toutes les PI3/4Ks mais sans conservation apparente entre famille de kinase. Par exemple, l'hélice C-terminale des PI4Ks et de VPS34 sont riches en résidus aromatiques alors que les autres PI3Ks en possèdent très peu. Chez les PI4Ks, cette hélice est strictement conservée chez les eucaryotes supérieurs et une conservation stricte du motif C-terminal 2099-IPY-2102 est observée entre PI4KA et STT4p. La délétion de cette hélice mène à une inactivation de l'enzyme dans le cas de PI4KB<sup>57</sup> et de VPS34<sup>58</sup>, c'est pourquoi il est essentiel de définir le rôle de cette hélice dans la catalyse. Enfin, cette hélice est amphipatique, comme l'hélice C-terminale de VPS34, et l'ouverture de l'hélice permet de découvrir sa surface hydrophobe.

Enfin, la **boucle de spécificité** est strictement conservée chez les PI4KA au sein des eucaryotes supérieurs, mais est variable entre PI4KA, STT4p et PI4KB. Seule la longueur de la boucle ainsi que le motif central 1984-PGG-1986 sont strictement conservés. Par conséquent, ce motif semble être le motif spécifique aux PI4Ks de type III. De manière surprenante, ce motif n'apparaît pas chez les PI3Ks phosphorylant le PtdIns. Cela peut être dû à la différence de positionnement de la kinase par rapport à son substrat qui ne phosphoryle pas le même groupement hydroxyl de l'inositol.

Seules deux structures cristallographiques de PI3Ks présentent une boucle de spécificité entièrement résolue : VPS34 en conformation « ouverte » (PDB 2X6F), ainsi que la PI3K $\beta$  (PDB 2Y3A) dans une conformation fermée et en interaction avec sa sous-unité régulatrice. Dans le cas de Vps34, le centre de la boucle de spécificité interagit avec la boucle d'activité et semble être repliée plutôt vers l'intérieur de la protéine. A contrario, la boucle de spécificité de la PI3K $\beta$  est étendue vers l'extérieur et interagit grâce à seulement deux résidus centraux avec sa sous-unité régulatrice de type  $\text{niSH2}$  inhibant son activité. Cette interaction avec des domaines de régulations pourrait expliquer la grande variation entre les boucles de spécificité de PI4KA et de PI3Ks phosphorylant le même substrat.

En amont de cette boucle de spécificité se trouve un motif conservé entre PI4Ks et PI3Ks, 1973-HIDFG-1977, comprenant l'aspartate 1975 impliqué dans la chélation d'un des magnésiums nécessaires à la réaction catalytique. Ce motif est d'autant plus important que la phénylalanine interagit avec les deux lobes du domaine catalytique. Le motif -DFG- est conservé aussi chez les sérine/thréonine kinases<sup>105</sup>, alors que -HI- est conservé au sein des PI4/3Ks et PIKKs.

Afin de tirer des informations supplémentaires de ces modèles, nous avons entrepris de réaliser un docking avec un inositol sur ces modèles ainsi que des simulations de dynamique moléculaire sur un modèle ouvert et fermé de la PI4KA humaine au préalable minimisés ainsi que sur certaines structures de PI3Ks. Cette étude serait le fruit d'une collaboration avec Thibault Tubiana et Yves Boulard.



**Figure 23-** Etude du domaine catalytique de la PI4KA : éléments principaux

**A-** Domaine catalytique de PI4KA modélisé sur les modèles VPS34 en forme ouverte (pdb 2X6H) et PI3K $\beta$  en forme fermée (pdb 2Y3A), grâce à I-tasser.

Eléments de PI3K $\beta$ : boucle de spécificité étendue vers l'extérieur en gris, hélice C-terminale et histidine catalytique dirigée vers l'ATP en bleu ; Eléments de Vps34 : boucle de spécificité en violet, histidine catalytique dirigée vers hélice C-terminale en jaune. A noter qu'il n'existe à ce jour aucun rapport connu entre le positionnement de la boucle de spécificité et celui de l'hélice C-terminale. Par ailleurs, la structure de PI3K $\beta$  a été résolue avec sa sous-unité régulatrice niSH2 et un analogue nucléotidique tout deux inhibiteurs de l'activité de la kinase. Enfin, L'ATP a été ajouté par superposition avec la PI3K $\gamma$  (pdb 1E8X)

**B-** Zoom sur la poche catalytique contenant un ATP et deux ions magnésium.

Cette figure a été réalisée à partir d'un de nos modèles « forme ouverte » obtenu avec I-tasser. Le modèle avec ATP et magnésium a été obtenu après superposition avec la seule structure de PI3K connue avec un ATP (PI3K $\gamma$ , pdb 1E8X) suivi d'une minimisation avec CHARMM27 sous GROMACS. Les résidus importants ont été modélisés en sticks, et colorés en blanc pour ceux appartenant à la boucle d'activité, violet pour la boucle de spécificité et rouge pour le lobe N-cat du domaine catalytique. Les deux sphères notées 2+ sont deux ions magnésium intervenant dans le transfert du phosphate  $\gamma$ .

## i) Matériel et méthode

### ▪ Alignement multiple par ClustalΩ

Les séquences utilisées pour l'alignement multiple ont été sélectionnées parmi les séquences annotées de PI4KA de vertébrés et d'insecte, ainsi que de STT4p de levure et de champignon de la banque UniProtKB (Uniprot Knowledge Base). Ces séquences ont été ensuite analysées par ClustalΩ, programme générant des alignements multiples par méthode progressive combinant deux approches : alignement de profils HMMs et construction d'arbre phylogénétique guide<sup>106</sup>. Les paramètres ont été utilisés par défaut et l'alignement a été analysé grâce à JalView<sup>107</sup>.

### ▪ Recherche d'homologues lointains : HHblits, HHPred et Phyre2<sup>108</sup>

L'une des difficultés a été de trouver des homologues, si possible de structure connue, de la partie N-terminale de PI4KA. Les alignements multiples de type PSI-Blast n'étant pas assez puissants, nous avons choisi d'utiliser HHblits en combinaison avec HHPred.

Dans un premier temps, des alignements multiples ont été générés par HHBlits à partir de la banque de séquences Uniprot20, avec 8 itérations. Cet alignement a été réalisé localement, et non globalement, pour une meilleure sensibilité. Afin d'affiner au maximum ces résultats, les analyses par HHBlits ont été réitérées 4 fois.

Dans un second temps, cet alignement multiple a été analysé par HHPred afin de trouver les homologues de structure connue dans la banque de données PDB la plus récente (pdb70\_18april15). Encore une fois, l'alignement par paire a été réalisé localement.

Un grand nombre de fragments de la séquence de la PI4KA humaines a été étudié, afin d'être certain de la délimitation des domaines et modules potentiels étudiés expérimentalement par la suite. Ils comprennent soit les domaines/modules présumés, soit des constructions plus grandes comprenant plusieurs domaines potentiels à la fois.

#### Liste des constructions testées :

<b><i>N-terminal</i></b>	1-1118, 1-1135, 1-DIII, 1-DI, DI-2102, DI-ARM, DIII-ARM
<b><i>Module 1-1100</i></b>	1-300, 1-553, 1-750, 20-300, 20-553, 20-750, 553-750, 553-1118, 750-1118, 300-750, 300-553
<b><i>Module DI-DIII</i></b>	DI-DIII, DI-DII, DII-DIII, DI, DII, DIII

Afin de compléter les recherches, la partie N-terminale de PI4KA a aussi été analysée par Phyre2 (**Protein Homology/AnalogY Recognition Engine**), un programme de modélisation par threading avec une stratégie un peu différente que celle d'I-tasser. En effet, Phyre2 recherche dans un premier temps des homologues par PSI-Blast à partir de la séquence de la protéine puis construit un profil HMM. Ce profil va être comparé à une banque de profils HMMs contenant seulement des séquences de protéines de la PDB. Ainsi, Phyre2 est très fiable dans la recherche d'homologues lointains. Les homologues détectés sont ensuite évalués.



Ce programme a été utilisé en mode intensif, ce qui permet de modéliser *ab initio* les fragments de la protéine pour lesquels aucun homologue n'a été trouvé en utilisant le programme Poing. Il s'avère que Phyre2 n'a généré des modèles que dans très peu de cas, et j'ai donc préféré l'utiliser plutôt comme détecteur d'homologue lointain de structure connue.

- Analogues structuraux : modélisation par threading

De nouveau, I-tasser a été utilisé mais cette fois-ci dans un but un peu différent. Il a été question de rechercher le type de repliement de chacun des domaines délimités auparavant, de s'assurer de leur autonomie, et si possible d'établir la fonction de chacun d'entre eux. En effet, n'ayant pas d'homologues de structure connue de la partie N-terminale de PI4KA, il a été question cette fois-ci de définir au mieux sa structure tridimensionnelle en se basant sur les meilleurs modèles générés ainsi que sur les analogues structuraux les plus proches. En effet, I-tasser utilise le logiciel TM align<sup>109</sup> afin d'aligner le meilleur modèle généré aux structures haute résolution répertoriées dans la PDB. L'alignement peut être évalué par le TM-score compris entre 0 et 1, pour lequel 1 est un alignement parfait. Un TM score inférieur à 0.2 correspond à un alignement random. Les paramètres d'I-tasser ont été utilisés par défaut, sans contraintes imposées.

- ii) Résultats

- Deux modules potentiels : 1-1100 et DI-2102

Au-delà des résultats fonctionnels obtenus par nos collaborateurs, les analyses bio-informatiques que nous avons réalisées nous ont permis de conclure que deux modules composent PI4KA : un module N-terminal, des résidus 1 à 1118, qui n'est pas essentiel à l'activité catalytique de PI4KA ; un module C-terminal, des résidus 1135 à 2102, comprenant en particulier un module potentiel de régulation DI-DIII. En effet, comme expliqué dans l'article précédent, une grande région désordonnée est prédite en amont de DI et en aval de DIII. De plus, la région désordonnée située en amont de DI est de longueur et de séquence variable. Ainsi, nous avons pu émettre l'hypothèse d'un module DI-DIII indépendant des 1100 premiers résidus de PI4KA. Au vu des résultats expérimentaux de nos collaborateurs, le module DI-DIII pourrait appartenir au module C-terminal de PI4KA.

Cette hypothèse a été confortée par les résultats obtenus lors des études bio-informatiques complémentaires. En effet, au cours des différentes analyses globales portant sur les constructions dites « N-terminal » (voir tableau p.73), nous avons pu détecter des homologues lointains de la région 300 à 1100 de la PI4KA humaine. En revanche, il n'en a pas été détecté pour le module DI-DIII, suggérant ainsi un repliement tri-dimensionnel différent des 1100 premiers résidus de la protéine. Ces données sont en accord avec les prédictions données par Interpro, méta serveur identifiant les domaines en se basant sur les relations évolutives<sup>110</sup>. Ce dernier détecte un domaine ARM seulement entre les résidus 78 à 1020 et 1530 à 1720 de la PI4KA humaine.

Nous avons décidé d'affiner nos recherches en étudiant séparément ces deux modules potentiels 1-1100 et DI-DIII pour lesquels la fonction et le repliement tri-dimensionnel restent inconnus.

- Etude du module DI-DIII

D'après nos alignements de séquences entre PI4KAs et STT4Ps, DI est le domaine putatif le mieux conservé des trois, suivi de DII qui est beaucoup plus variable. DIII est quant à lui très variable, seule une hélice est assez bien conservée en amont du domaine ARM.

Les recherches d'homologues lointains n'ayant pas abouti, j'ai entrepris d'analyser essentiellement les prédictions de structures tridimensionnelles du module DI-DIII.

Nous savons d'ores-et-déjà, grâce aux prédictions de structure secondaires, que le module DI-DIII est prédit quasi entièrement hélical et que la limite entre les domaines putatifs est marquée par des régions prédites désordonnées (voir article n°1). Cependant, la limite entre DII et DIII est une région prédite désordonnée étendue sur une centaine de résidus (1344-1478) englobant environ la moitié du domaine putatif DIII. A noter que cette boucle potentielle comprend un résidu phosphorylable, la sérine 1436<sup>111</sup>.

Tout comme les prédictions de structures secondaires, le serveur I-tasser démontre aussi la présence de trois domaines putatifs distincts. En effet, l'étude de DI-DIII a nécessité trois types de templates distincts pour chacun des domaines, soit trois types de repliements tridimensionnels différents. Par ailleurs, le C-score permettant l'évaluation du meilleur modèle généré (C-score de -2,22 ; Tm 0,45) ainsi que le Tm de l'analogue structural le plus proche (0,617) étant moyens, j'ai entrepris d'étudier les domaines DI, DII et DIII séparément avec I-tasser et Phyre2.

Le résultat le plus fiable obtenu avec I-tasser est le meilleur modèle obtenu du domaine putatif DI exposé en figure 24. Il possède un C-score de -1.9 et un Tm de 0.49. Le modèle est composé de 6 hélices  $\alpha$  formant un domaine assez compact. Les deux analogues structuraux les plus proches, avec un TM de 0.798 et 0.735 respectivement, sont des domaines de liaison à l'ADN de protéines régulatrices de la transcription bactérienne (PDB 3IHU, 3C7J). Tout comme le meilleur modèle I-tasser, ils adoptent un repliement assez compact, formé de 6 hélices agencées en clé grecque (repliement « GNTR ligand-binding domain like »).

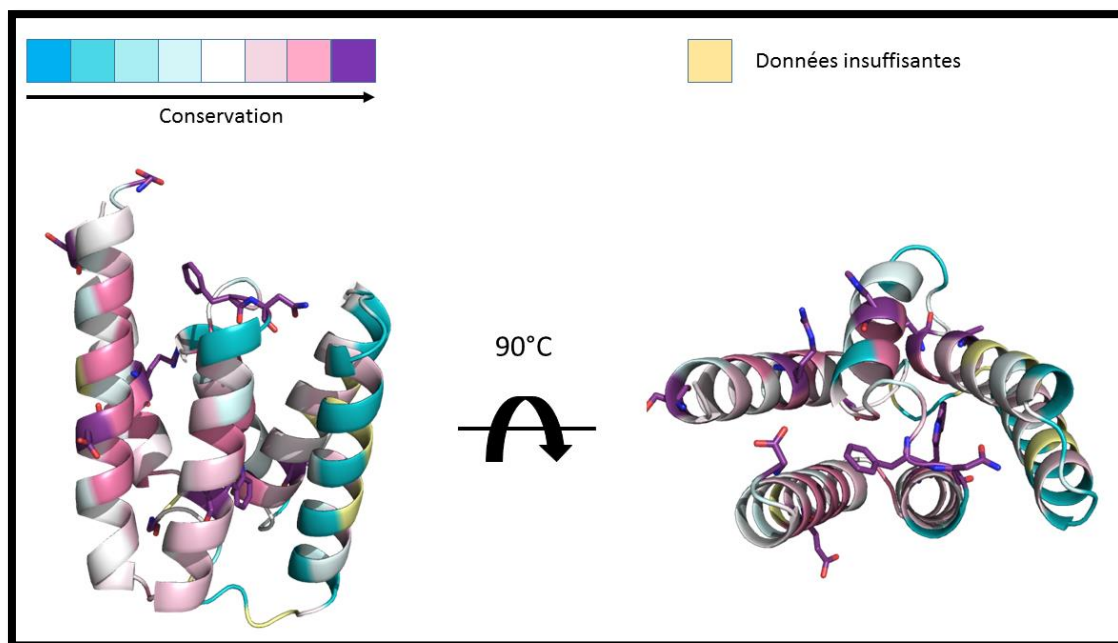
Le modèle obtenu pour le domaine putatif DII est moins fiable, mais reste dans les normes (C-score -3.3, Tm 0.35). L'analogue structural le plus proche (Tm 0.712, pdb 1Z3H), est une importine de levure qui médie l'export de l'importine  $\alpha$  du noyau au cytoplasme. Comme DI, il est formé de 6 hélices, cependant elles sont agencées en répétitions HEAT et adopte donc un repliement en super-hélice  $\alpha$ , formant ainsi un feuillet d'hélices incurvé.

Malgré des repliements globaux différents, plusieurs éléments semblent montrer que les domaines DI et DII appartiennent au même module :

- Les analogues structuraux les plus proches de DI et DII appartiennent à des règnes « inférieurs » : la levure et la bactérie. Ainsi, la relation séquence-repliement semble n'être retrouvée qu'au sein de ces organismes, alors que les modules N- et C-terminaux les entourant sont analogues à des protéines humaines.
- Les protéines analogues les plus proches interagissent toutes avec des acides nucléiques, soit pour leur activité, soit pour réguler des assemblages de complexes protéiques. De plus, DI possède un C-score de 0.68, faible mais non négligeable, en ce qui concerne l'interaction entre DI et les acides nucléiques. Même s'il n'est pas rare que ces prédictions soient confondues avec des interactions aux phospholipides, il n'en reste pas moins que DI-DII pourrait interagir avec un cofacteur, expliquant ainsi leur hypothétique rôle dans la régulation de l'activité de PI4KA.

Ainsi, nous pouvons supposer que DI-DII soit un module de régulation de l'activité catalytique de PI4KA, voire même un domaine autonome liant les modules N- et C-terminaux. Cette régulation pourrait se faire par l'intermédiaire d'interaction avec un acide nucléique ou avec la tête polaire du PtdIns ainsi que par la phosphorylation de la sérine 1436.

Enfin, en ce qui concerne le domaine putatif DIII, I-tasser n'a pas généré de résultats fiables. Au vu de la grande région désordonnée composant sa partie N-terminale ainsi que sa non-conservation entre les espèces, il se pourrait que l'hélice conservée de DIII fasse le lien entre le domaine ARM et le module DI-DII.



**Figure 24-** Meilleur modèle de DI obtenu avec I-tasser.

L'alignement de séquence entre PI4KAs et STT4ps en annexe 2 a été utilisé en combinaison avec Consurf, programme permettant de colorer les résidus du modèle par degré de conservation<sup>112</sup>. Globalement, le domaine putatif DI n'est pas très conservé.

- Etude du module N-terminal 1-1118

Une grande région désordonnée après les résidus 1118, retrouvée dans les prédictions de structure secondaires de QUICK2D, Phyre2 et I-tasser, suggère que le module N-terminal se termine aux alentours du résidu 1118. Les études menées sur ce module portent ainsi sur la séquence des résidus 1 à 1118 de la PI4KA humaine.

Les résidus N-terminaux de la PI4KA humaine ne sont pas conservés entre STT4p et PI4KA. Cependant, ils semblent assez bien conservés entre vertébrés. En l'absence de structure tridimensionnelle, peu d'éléments sont exploitables dans ces alignements de séquences, cependant deux détails sont frappants :

- La présence de **poly-glycines** (résidus 1-24)

La méthionine N-terminale n'ayant été définie que récemment<sup>50</sup>, les premiers résidus des différentes séquences de PI4KA sont souvent manquants. Il s'avère que les résidus 1 à 24 de la PI4KA humaine sont fortement enrichis en glycines. Ces glycines peuvent être remplacées, dans certains cas, par des alanines dans le cas de la poule, voire des séquences sans répétitions en acides aminés pour les insectes et les eucaryotes inférieurs. Deux hypothèses s'offrent à nous :

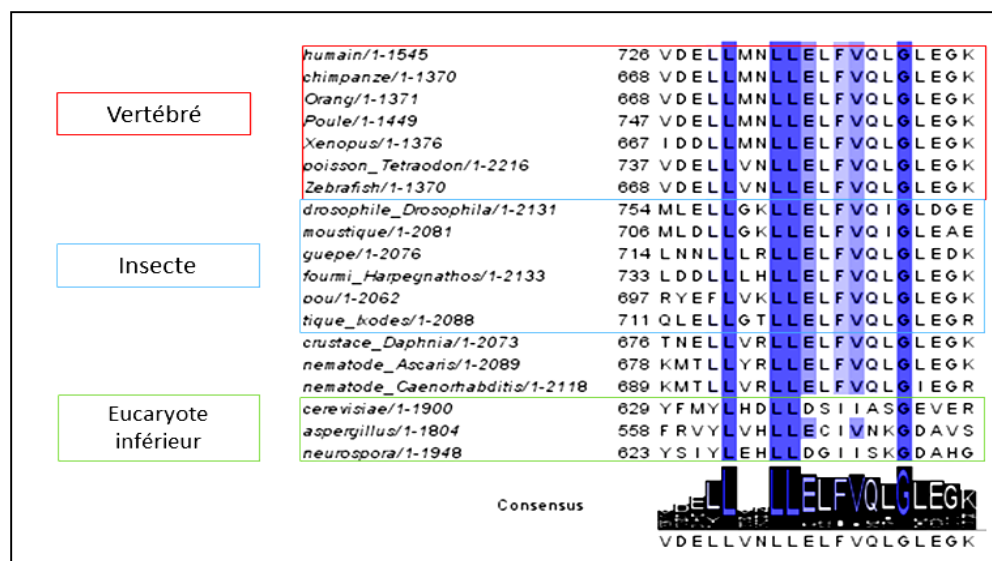
- La séquence en acide nucléique correspondante est riche en GC. Elle peut donc réguler la transcription de la PI4KA grâce à la formation de hairpin
- Ce type de séquence est retrouvé par HHblits/HHpred dans un certain nombre de protéines possédant un peptide signal, dans notre cas peut être spécifique aux vertébrés. De plus, Prosite prédit que les 20 premiers résidus N-terminaux constituent un site d'attachement aux membranes chez les lipoprotéines procaryotes.

Ainsi, que ce soit au niveau de la séquence en acides nucléiques ou en acides aminés, les 20 premiers résidus de PI4KA seraient importants dans la régulation de la protéine.

- La région 729-**LxxLLELFVQLGL**-742, **riche en leucines**

Cette dizaine de résidus est strictement conservée entre PIK4As, et les trois premières leucines ainsi que le résidu acide sont conservés entre PI4KA et STT4p. Sa séquence est riche en résidus hydrophobes, en particulier riche en leucines. D'après les prédictions de structures secondaires, il s'agirait du C-terminal d'une hélice juste en amont de deux hélices prédites trans-membranaires par un certain nombre de programmes (TM HMMTOP<sup>113</sup>, TM Phobius<sup>114</sup>, TM Pred).

Or, PI4KA est connue pour être une protéine périphérique aux membranes, sans ancrages à cette dernière. Cependant, au vu de la taille de la protéine, ces hélices peuvent être tout simplement des hélices enfouies au sein de PI4KA.



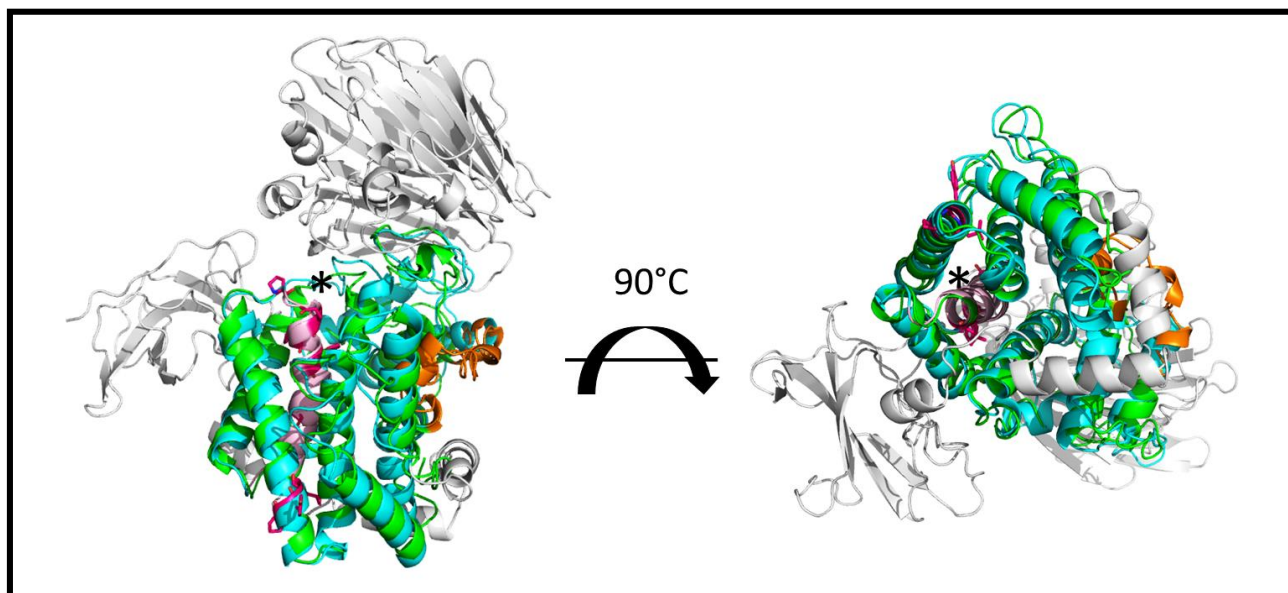
**Figure 25-** Alignement de séquence du motif consensus riche en leucines.

Sur la base des prédictions de structure secondaire, et malgré la prédiction d'un module N-terminal quasi entièrement hélical, plusieurs domaines potentiels ont pu être définis :

Le tout premier domaine potentiel, situé entre les résidus **24 à 300**, est non conservé même entre vertébrés et est beaucoup plus court chez les eucaryotes inférieurs. Lors des analyses du segment 1-1118, aucun homologue lointain n'a été détecté pour les résidus 24-300 au contraire de la région 300-1118. Par conséquent, j'ai émis l'hypothèse que les résidus 24-300 possédaient un repliement autonome et pouvaient être un domaine à part entière.

Dans les résidus 255 à 279 est prédite une région désordonnée riche en sérine (11% des résidus chez la PI4KA humaine) et en prolines. Parmi elles, la sérine 265 est phosphorylable, et les sérines 199, 204 et 207 ont été identifiées comme phosphorylables lors d'analyses spectrométriques massives<sup>115</sup>. Bien que variable, une telle boucle riche en sérine est observable chez les vertébrés et les insectes, mais pas chez les eucaryotes inférieurs. De plus, ce domaine est prédit quasi entièrement hélical, avec des brins bêta potentiels pour lesquels les prédictions sont difficilement interprétables. En effet, le nombre et la longueur des brins et des hélices prédits par Quick2d, Phyre2 et I-tasser sont différents. La difficulté de prédiction de la structure de ce fragment est aussi détectable sur le meilleur modèle généré par I-tasser, dont le C-score et le Tm sont un peu faibles (C-score de -3,08, Tm de 0.37). Ce modèle présente un repliement entièrement hélical et compact. Les hélices s'agencent de la même manière que l'analogue structural le plus proche : le domaine N-terminal d'alginate lyase bactérienne avec un repliement de type chondroïtin AC/alginate lyase (PDB 3A0O, Tm 0.905 très fiable).

Au vu des prédictions assez contradictoires, nous ne pouvons pas nous fier au repliement tri-dimensionnel proposé. Des données structurales sont requises.



**Figure 26-** Modèle du domaine putatif 20-300 (vert) superposé à la structure de l'alginate lyase d'*Agrobacterium Tumefaciens* (pdb 3A0O, coloré en cyan et gris)

La boucle riche en sérine et proline est colorée en orange, et est prédite accessible au solvant par I-tasser. A noter que la seule hélice conservée de ce domaine potentiel (notée (\*)) et colorée en rose est enfouie. Les résidus strictement conservés chez les eucaryotes supérieurs (en rose foncé) sont essentiellement des résidus hydrophobes, dont deux phénylalanines ainsi qu'un glutamate conservé entre STT4p et PI4KAs. Ainsi, les alignements de séquence et ce modèle sont en accord quant à une place charnière de cette hélice dans le repliement du domaine 20-300.

Le second domaine potentiel est situé entre les résidus **445 à 1118**.

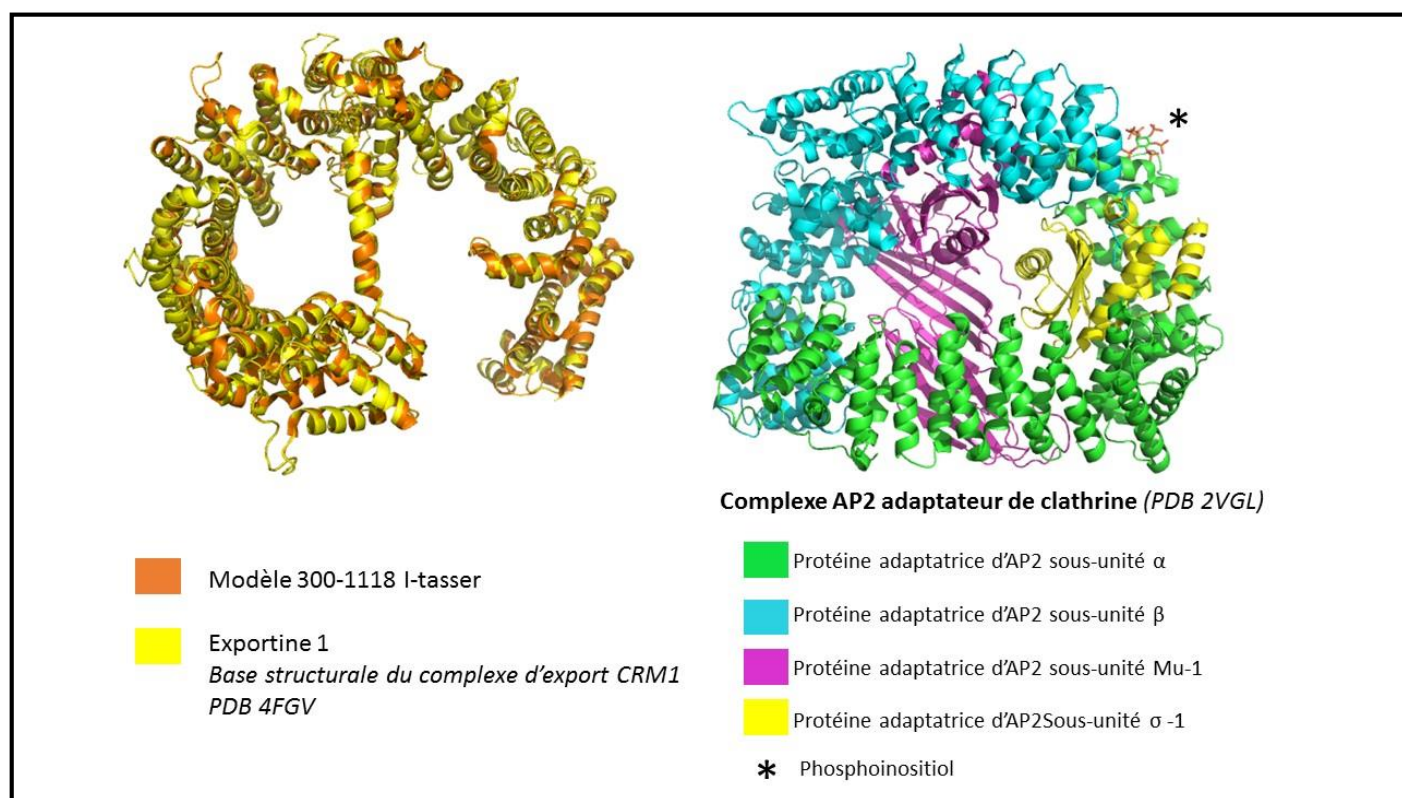
Les homologues de cette région adoptent tous le même type de repliement en répétition ARM, suggérant ainsi une fonction dans le recrutement de partenaires protéiques. L'homologue le plus fiable (Phyre2, 95.6% ; HHPred : 95.3%) est détecté pour les résidus 445 à 737 de la PI4KA humaine. Il s'agit de la sous-unité  $\beta$  de la protéine AP2, une protéine adaptatrice de clathrine (pdb 2VGL\_B). Comme expliqué précédemment (Intro.II.4), cette protéine intervient dans la formation d'un complexe protéique permettant la formation de vésicule mantelée et le recrutement de cargo. De manière intéressante, ce complexe est activé par le PtdIns(4)P généré par la PI4KII $\alpha$  au sein du Golgi. La sous-unité  $\beta$  d'AP2 présente un repliement constitué de répétitions HEAT, formant ainsi un feuillet d'hélice avec une forte courbure (Figure 27). Le meilleur modèle obtenu (C-score -2.13, TM 0.46) présente lui aussi un repliement hélical de type répétitions HEAT, tout comme la sous-unité  $\beta$  d'AP2. Cependant, l'analogue structural le plus proche (PDB 4fgv\_A, Tm 0.793) correspond cette fois-ci à la partie N-terminale de l'exportine 1, une protéine de transport de macromolécules au noyau. Ce domaine de la protéine correspond à la base structurale du complexe, tout comme la sous-unité  $\beta$  d'AP2 dans le complexe présenté figure 27. Cette région de la protéine, tout comme notre modèle,



appartient à une superfamille différente de celle d'AP2, du type importine-exportine  $\beta$  like. Cependant, elle a la même fonction globale que la sous unité  $\beta$  d'AP2 et un type de repliement très similaire.

Pour chacune de ces études, le repliement global est donc le même : entièrement hélical et composé de répétitions HEAT. De plus, il semblerait que la structure tridimensionnelle corresponde au domaine de protéines intervenant dans le transport de protéine. Or, les répétitions HEAT sont connues pour intervenir dans le recrutement de partenaires protéiques. Le module N-terminal de PI4KA pourrait ainsi être impliqué dans le transport de la protéine du RE vers la membrane plasmique en interagissant entre autre avec TTC7.

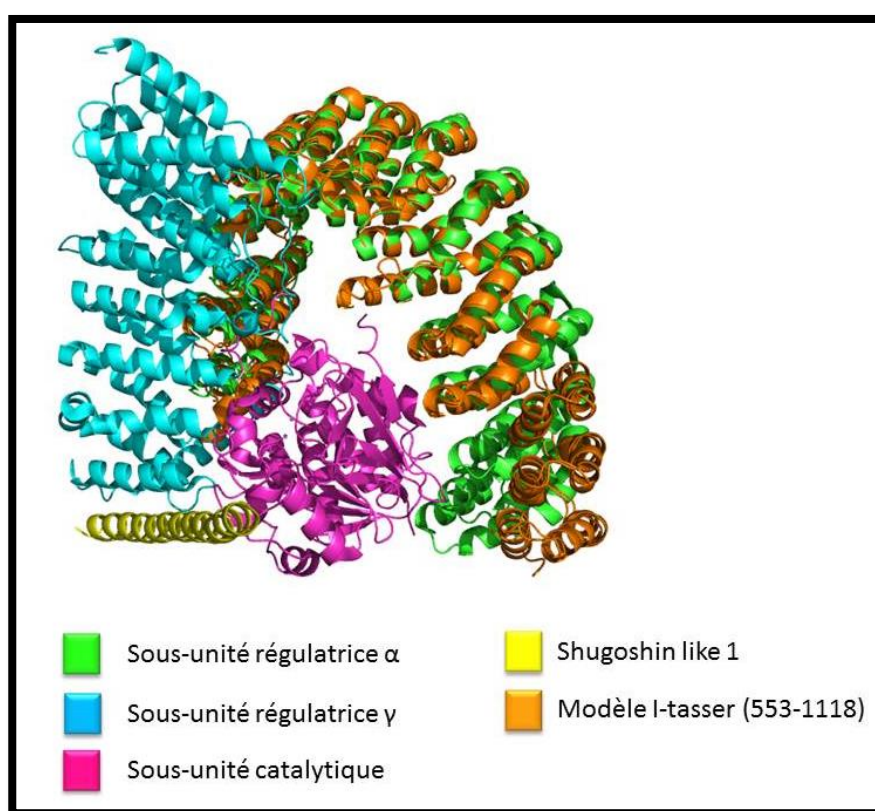
Enfin, au vu de la longueur du domaine potentiel, cette répétition d'hélice est assez étendue pour former un fer à cheval. De telles structures, dans les protéines de transport, sont connues pour entourer en partie la macromolécule à transporter. Ainsi, le domaine catalytique pourrait venir se loger au sein de ce module N-terminal 1-1118.



**Figure 27-** Modèle obtenu par I-tasser superposé à son analogue structural l'exportine 1 et comparaison avec le complexe d'AP2, complexe adaptateur de clathrine.

Le modèle généré, l'exportine 1 et la sous-unité  $\beta$  du complexe AP2 homologue à PI4KA sont toutes trois formées de répétitions HEAT. Leur structure globale est fortement courbée et, dans le cas d'AP2, permettent le recouvrement de la protéine à transporter.

Néanmoins, les prédictions ne sont pas en accord quant à la délimitation N-terminale du domaine à répétition HEAT. En effet, les régions prédites désordonnées sont prédites des résidus 255 à 279 et de 545 à 553. Il s'avère que les résultats générés pour la construction 553-1118 sont les meilleurs obtenus au sein des analyses du module N-terminal. Le modèle généré est assez fiable (C-score -1.47, TM 0.53) et son analogue structural le plus proche, loin devant les autres, a été évalué avec un TM de 0.914 et un RMSD de 1.34. Le modèle est donc très proche de son analogue structural qui est par ailleurs aussi détecté par HHPred. Cet homologue est le domaine régulateur de la sérine/thréonine phosphatase PP2A humaine (pdb 1B3U), composé de 15 répétitions HEAT, ce qui est en accord avec les prédictions concernant le segment 445-1118 de PI4KA. De plus, cette protéine coordonne l'assemblage de la sous-unité catalytique de la PP2A avec sa sous-unité régulatrice  $\gamma$  (figure 28).



**Figure 28-** *Complexe de la sérine/thréonine protéine phosphatase PP2A avec l'un de leur substrat, la protéine shugoshin-like 1 (PDB 3FGA).*

Ce complexe a été superposé avec le modèle structural du fragment 553-1118 de la PI4KA humaine. A noter que la sous-unité catalytique de PP2A se trouve au sein du fer à cheval formé avec l'homologue du domaine putatif 553-1118, la sous-unité régulatrice  $\alpha$  de la PP2A. De plus, la sous-unité régulatrice  $\gamma$  interagit d'une part avec le domaine à répétitions HEAT grâce à sa grande hélice C-terminale ; et d'autre part avec la sous-unité catalytique de PP2A grâce à une longue boucle.



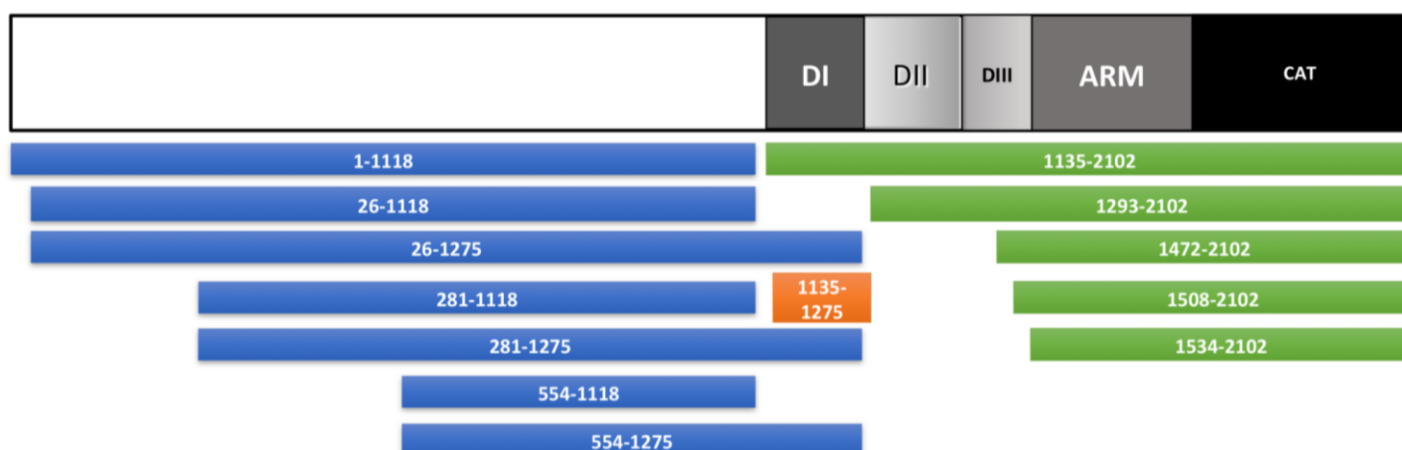
En conclusion, nous avons pu proposer le découpage du module N-terminal suivant :

- Les résidus 1 à 24, intervenant potentiellement dans la régulation de PI4KA (localisation, traduction ?)
- 28-276, domaine hélical et compact qui se termine par une région désordonnée avec des sérines phosphorylables
- X-1118, module replié en répétitions HEAT, permettant certainement le recrutement de partenaires ainsi que le transport de la PI4KA en logeant peut-être le domaine catalytique au sein même du module N-terminal susceptible de prendre la forme d'un fer à cheval. Ce domaine putatif possède un motif riche en leucine très conservé et est composé potentiellement de deux sous-domaines dont l'un débiterait vers le résidu 553

## B. Délinéation expérimentale des modules et domaines composant PI4KA

### 1) *Article n°2: Definition, expression in E.Coli and purification of large fragments from the human lipid kinase PI4KA*

La purification de PI4KA entière en baculovirus ne permettant pas d'obtenir suffisamment de matériel afin de réaliser des caractérisations biophysiques (partie III), nous avons décidé d'exprimer, purifier et caractériser des fragments de PI4KA en bactérie. Pour cela, il a fallu dans un premier temps définir des constructions solubles. Par conséquent, nous nous sommes basés sur notre découpage de la PI4KA humaine exposée dans le chapitre précédent grâce à nos études bio-informatiques. Voici les différentes constructions testées expérimentalement :



Toutes ces constructions ont été produites en fusion avec une étiquette GST en système bactérien *E.Coli* BL21(DE3) C41. Elles ont toutes été exprimées et purifiées dans les mêmes conditions afin de pouvoir comparer le comportement en solution de chacune d'entre elles. Les protéines de fusion ont été purifiées en batch grâce à de la résine de glutathion. Les étapes de lyses et de purification ont été réalisées dans un tampon faible en sel (100Mm NaCl) et sans détergent afin d'évaluer au mieux leur solubilité.

Nous avons pu ainsi mettre en évidence différents points :

- Les constructions C-terminales débutant à DIII ou plus loin sont purifiées en présence d'une grande quantité de chaperonnes, attestant de leur mauvais repliement.
- Les constructions 1135-2102 et 1293-2102 se comportent beaucoup mieux, mais sont beaucoup moins exprimées, sûrement à cause de leur grande taille et de leur repliement complexe. Ces constructions seront étudiées plus en détail dans le chapitre suivant. Ces résultats confortent l'idée d'un module DI (1135-1275) avec un repliement autonome, voire un domaine pivot entre les modules N- et C-terminaux.
- Deux familles de constructions N-terminales ont été testées : x-1118, sans le domaine DI ; et x-1275, en présence de DI. Ainsi, nous avons voulu nous assurer que le domaine putatif DI appartenait au module C-terminal. Il s'avère que les constructions avec DI ne sont pas solubles, confortant ainsi notre hypothèse
- Les constructions 26-1118 et 554-1118 sont les constructions avec l'expression et la solubilité les plus élevées. Par conséquent, le module N-terminal semble bien être composé de deux domaines.
- La construction 281-1118 n'est purifiée quant à elle qu'en faible quantité. La délimitation du domaine potentiel était incertaine au vu des prédictions de structures secondaires contradictoires. Nous ne pouvons donc pas tirer de conclusion quant à l'existence d'un domaine N-terminal 1-300.

Nous avons décidé d'étudier la structure tri-dimensionnelle de la construction 26-1118 afin de tirer le plus d'informations possible sur le module N-terminal de la PI4KA dans son entier. De plus, nous avons tenu à garder les sites potentiels d'interaction avec le domaine d1 de la protéine virale NS5A. Par conséquent, nous avons mis au point l'expression et la purification en quantité suffisante du module 26-1118, protocole explicité dans l'article qui suit.



Contents lists available at ScienceDirect

## Protein Expression and Purification

journal homepage: [www.elsevier.com/locate/yprep](http://www.elsevier.com/locate/yprep)

# Definition and expression in *E. coli* of large fragments from the human lipid kinase phosphatidylinositol 4-kinase type III alpha, and purification of a 1100-residue N-terminal module



Cyntia Taveneau, Karine Blondeau, Stéphane Bressanelli\*

Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Université Paris-Sud, CEA Saclay, 91191 Gif-sur-Yvette cedex, France

## ARTICLE INFO

## Article history:

Received 16 June 2015

Accepted 29 June 2015

Available online 2 July 2015

## Keywords:

Phosphatidylinositol 4-kinase

PI4KA regulation

Bio-informatics

Production in *E. coli* of large recombinant proteins

## ABSTRACT

The eukaryotic lipid kinase phosphatidylinositol 4-kinase III alpha (PI4KA in higher eukaryotes) is a ubiquitous enzyme that synthesizes the plasma membrane pool of phosphatidylinositol 4-phosphate. This important phosphoinositide has key roles in different signalization pathways, vesicular traffic and cellular compartment identity. Moreover, human PI4KA is an essential factor for hepatitis C virus replication. PI4KA is a large protein (2102 residues for human PI4KA) with the kinase domain making up the ca 400 C-terminal residues. There is essentially no structural information about the 1500 N-terminal residues and no clue as to the function of most of this region of PI4KA.

In this report, we use computational methods in order to delineate fragments of human PI4KA amenable to soluble production in *Escherichia coli*. We clone and express these fragments as GST-fusions and evaluate the soluble fraction of each protein. Finally, we produce and purify to homogeneity a 1100-residue PI4KA N-terminal fragment. Our results further suggest that PI4KA can be described as a two-module protein. They open the way to structural characterization of the N-terminal regulatory module of PI4KA.

© 2015 Elsevier Inc. All rights reserved.

## 1. Introduction

Phosphatidylinositol-4-kinases are essential eukaryotic lipid kinases synthesizing phosphatidylinositol-4 phosphate (PI4P) by phosphorylating phosphatidylinositol at the D-4 position of the inositol ring. PI4P is an important phosphoinositide with a key role in different signalization pathways, vesicular traffic and cellular compartment identity [1,2]. Several distinct families of phosphatidylinositol-4-kinases are known. The so-called type III enzymes (hereafter called PI4KIII) were defined on their inhibitor sensitivity profile and molecular weight and are known to be homologous to the phosphatidylinositol-3-kinase (PI3K) family [1–3]. The C-terminal region is conserved throughout the eukaryote kingdom across PI4KIIIs and PI3Ks, especially the catalytic domain with about 27% pairwise sequence identity between the

closest PI4KIII and PI3K [4]. In contrast, there is no detectable sequence identity between N-termini. Indeed, N-termini are extremely variable both in length and domain composition among PI4KIIIs/PI3Ks (see below).

In human cells, two distinct PI4KIIIs have been identified [2]. The first, PI4K230 or PI4KA, is a large protein of 2102 residues (the shorter isoform that was also reported is most probably non-functional) that is mainly located at the periphery of the endoplasmic reticulum. The second, PI4K92 or PI4KB, comprises 801 residues and is localized at the Golgi compartment. In accordance with PI4KA and PI4KB's role in membrane compartment identity, single-stranded, positive-sense RNA viruses, that depend on rearrangement of intracellular membranes for replication, have been reported to recruit one or both of these enzymes [5–7]. For instance, PI4KA is an essential factor in the formation of the endoplasmic reticulum-derived “membranous web” where hepatitis C virus (HCV) replication takes place. PI4KA's interaction with the non-structural HCV protein NS5A leads to the accumulation of PI4P in the membranous web giving to the membrane the signature necessary to the formation of viral replication machinery [6,8].

The C-terminal kinase domain of about 400 residues is highly conserved in PI4KA orthologs (52% identity with *Drosophila*

**Abbreviations:** PI4KIII, phosphatidylinositol-4-kinase of type III; PI4KA, (human) phosphatidylinositol-4-kinase III $\alpha$ ; PI4P, phosphatidylinositol-4 phosphate; PI3K, phosphatidylinositol-3-kinase; HCV, hepatitis C virus; SEC, size exclusion chromatography.

\* Corresponding author at: I2BC, Bât 532, CEA Saclay, 91191 Gif-sur-Yvette cedex, France.

E-mail address: [stephane.bressanelli@i2bc.paris-saclay.fr](mailto:stephane.bressanelli@i2bc.paris-saclay.fr) (S. Bressanelli).

*melanogaster*, 41% with *Saccharomyces cerevisiae*) and paralogs (50% with human PI4KB) [4]. In all PI4Ks and PI3Ks, this catalytic domain is preceded by an  $\alpha$ -helical repeat domain with an armadillo-type (ARM) fold, that is often called the lipid kinase unique (LKU) or PI3K accessory (PIK) domain. The ARM domains of PI4Ks/PI3Ks greatly differ in sequence and size, with various numbers of helical repeats, and are involved in scaffolding and protein–protein interactions [9,10]. N-terminal to the ARM domain can be further regulatory domains involved in protein–protein and protein–lipid interactions that contribute to recruitment and activation at intracellular membranes. For instance, in the catalytic subunit of the class IA PI3K there are three mostly  $\beta$ -sheet domains N-terminal to the  $\alpha$ -helical ARM domain [11]. Similarly for PI4KA, several studies show involvement of the N-terminal region (ARM and upstream regulatory domains) in subcellular localization and activation. Thus, PI4KA is dynamically recruited to the plasma membrane through its 58 N-terminal residues including the glycine-rich N-terminus [12]. This resolves the apparent paradox that PI4KA provides the plasma membrane with PI4P but appears located at the endoplasmic reticulum [1,2]. Recently, it was shown that HCV also recruits PI4KA by targeting regions well upstream of the catalytic domain [13,14], further highlighting the importance of the PI4KA N-terminal region in localization and activity regulation.

In contrast to the C-terminus of PI4KA (catalytic and part of ARM domains), for which structures are available for homologous PI3Ks [15] and now PI4KB [10], no information is available about the structures or even boundaries of domains in the N-terminus of PI4KA (residues 1–1530). The delimitation of domains in this N-terminal region of PI4KA is difficult because of (i) its large size (ii) the lack of detectable homologs of known structure (iii) the fact that it is reliably predicted as almost all  $\alpha$ -helical. Indeed, it is not clear *prima facie* where the PI4KA ARM domain begins. In this manuscript, we define large fragments of PI4KA that are amenable to soluble expression in *Escherichia coli* and produce and purify a biochemically well-behaved 1100-residue putative N-terminal module.

## 2. Material and methods

### 2.1. Construct design

Protein constructs representing putative domains of PI4KA were designed as follows: A variety of bioinformatics analyses were performed to delineate possible boundaries of domains within the PI4KA sequence using the MPI Bioinformatics Toolkit web service [16]. Secondary structure and disorder predictions were obtained

through QUICK2D: a multiple sequence alignment is generated with PSI-BLAST (up to 8 iterations) and automatically forwarded to PSIPRED, JNET and Prof. for secondary structure predictions and to IUPRED for disorder predictions.

### 2.2. Cloning of constructs into the pGEX-6p-1 expression vector

We amplified by PCR DNA sequences corresponding to the different constructs from the full-length PI4KA gene sequence (Gene Bank accession no: AB384703) contained in a recombinant pFastBacDual plasmid. The primers used for PCR are described in Table 1.

DNA amplifications were performed in a total volume of 50  $\mu$ L containing 60 ng of template, 0.5  $\mu$ M of each primer, 0.4 mM of deoxynucleotide triphosphate mix (Biolabs New England), 1 $\times$  PCR buffer containing  $MgSO_4$  and at last 2.5 units of PWO polymerase from the kit PWO DNA Polymerase (Roche).

The following PCR program was used for all needed DNA amplification: hot start at 94 °C for 5 min, followed by 30 cycles of denaturation at 94 °C for 30 s, annealing at 69 °C for 30 s, and extension at 72 °C. The time of extension was adapted to the size of the sequence according to the protocol provided by Roche. The program finally ends with an extension at 72 °C for five minutes. The PCR products were analyzed in 0.8% agarose gel and purified with the kit Nucleospin Gel and PCR clean-up (Macherey Nagel).

PCR products and pGex-6p-1 were digested with the appropriate restriction endonucleases (Table 1) and were ligated to each other with the Rapid DNA Ligation kit from Roche, with a vector:insert ratio equal to 1:5. Ligation product was then transformed into *E. coli* DH5 $\alpha$  using the heat shock method. Colonies were analyzed by restriction enzyme digestion and plasmids were purified from positive colonies. All PI4KA inserts were sequenced prior to expression tests (GATC Biotech, Germany).

### 2.3. Small scale expression analysis

#### 2.3.1. Expression

The bacterium strain used for protein expression was *E. coli* BL21(DE3)C41. After transformation of the recombinant plasmids, a single colony was subcultured at 37 °C under agitation in 5 mL Luria Broth (LB) medium containing 100  $\mu$ g/mL ampicillin. The next day, this subculture was diluted in 30 mL of fresh medium with appropriate antibiotics to a final optical density about 0.1 at a wavelength of 600 nm ( $OD_{600}$ ). Cells were cultivated at 37 °C with shaking until they reached an  $OD_{600}$  about 0.8. Protein expression was then induced at 18 °C with 1 mM isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG). Cells were then

**Table 1**  
Primers used to PCR clone the various constructs.

N-terminal residue number	Forward primer	Restriction enzyme site	C-terminal residue number	Reverse primer	Restriction enzyme site
<i>N-terminal constructions</i>					
1	gcggaattcgccgccccggccggggagg	EcoRI	1118	ggcgtcgacttatgtgttctgtttagcc	Sall
26	gcggaattcgctcgcgggcttctatt	EcoRI	* 1118	ggcgaattcttatgtgttctgtttagcc	EcoRI
281	gcggaattccctggagatctgccttc	EcoRI	1275	ggcgtcgacttagtctgttcttattctcagc	Sall
554	gcgggatcatgtcgggtaagaagacca	BamHI			
<i>C-terminal constructions</i>					
1135	gcgggatccgactactccaactcatggcatc	BamHI	2102	ggcgaattctaatagggatgtcattct	EcoRI
1293	gcgggatccacccccactacatctggat	BamHI			
1472	gcgggatccgcatatgagcctgagtgaagcagtg	BamHI			
1508	ggcggatccaacccgtgtcagccc	BamHI			
1534	gcgggatccagctgagtgaagcagtg	BamHI			

The restriction sites introduced are indicated as well as the resulting PI4KA N-terminal residue number (forward primers) and C-terminal residue number (reverse primers).

\* Primer used only for construct 554–1118.



incubated overnight and finally harvested by centrifugation 20 min at 3500×g. Pellets were stored at −20 °C.

### 2.3.2. Pulldown of soluble GST-tagged material

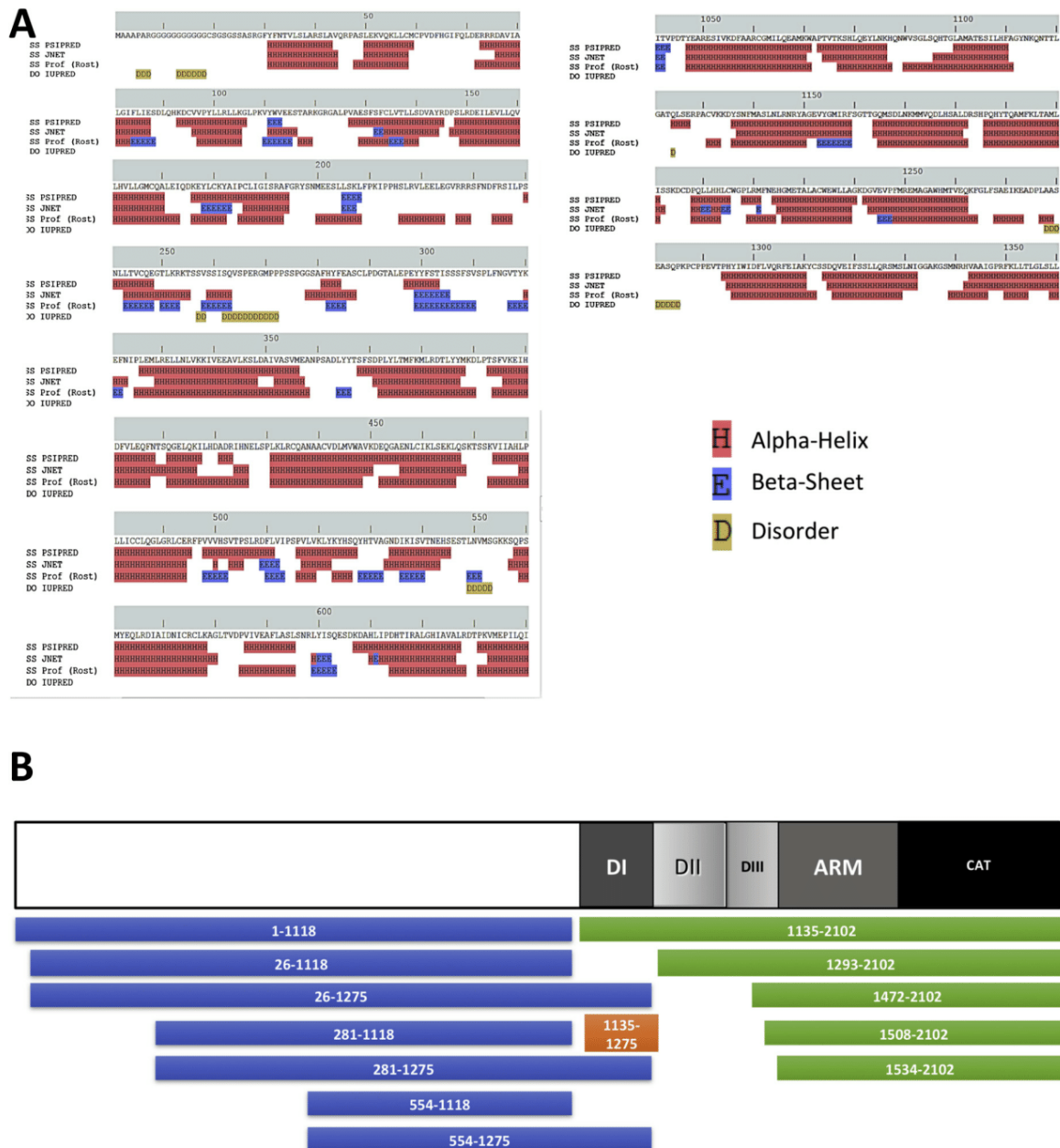
The entire process was carried out at 4 °C. Pellets were resuspended in 1 mL of lysis buffer composed of 25 mM Tris–HCl pH 7.4, 100 mM NaCl, 10% glycerol. Cells were then sonicated with 8 pulses of 3 s at 40% amplitude. The lysate was centrifuged for 20 min at 20,000×g and the supernatant was incubated with 20 µL of Protino Glutathione Agarose 4B beads (Macherey Nagel) under agitation for an hour. Beads were washed 3 times with 800 µL of lysis buffer and finally were heated for 5 min at 90 °C with 5 µL of Laemmli sample buffer 5× in order to denature proteins. Samples were separated on 14% polyacrylamide SDS–PAGE.

Pulled down proteins were identified by MALDI-TOF mass spectrometry after excising bands and digesting them with trypsin.

### 2.4. Large scale production and purification

#### 2.4.1. Production

GST-tagged PI4KA fragment 26–1118 was produced in an Applikon fermentor system in batch cultures of the *E. coli* strain BL21(DE3)C41. After transformation with the recombinant pGex-6p-1 a single colony was subcultured first into 5 mL and then into 500 mL of fresh 2YT medium supplemented with 100 µg/mL ampicillin at 37 °C under agitation. 2 parallel cultures of 5 L were carried out in LB medium at 37 °C with 100 µg/mL ampicillin, initial OD<sub>600</sub> of 0.1, initial pH of 7.0, 600 rev min<sup>−1</sup> and 1 vvm



**Fig. 1.** Sequence analyses and delineation of constructs. (A) Secondary structure and disorder predictions in the N-terminus of PI4KA. Screen shot of the Quick2D results obtained for the first 1360 residues of PI4KA. The sequence of the protein is displayed below the gray scale with the numbering of the residues. (B) Map of PI4KA and the constructs tried for expression and purification in fusion with GST.



(one volume of air per volume of culture per minute). Cells were grown up to an OD<sub>600</sub> of 0.8. Temperature was then brought to 23 °C and protein expression was induced with isopropyl-β-D-thiogalactopyranoside 1 mM for 15 h, *i.e.*, one additional cell generation. Cells from the 2 cultures were collected by centrifugation (12 × 800 mL, 5000×g, 20 min at 8 °C). Each of the 12 pellets was resuspended in 30 mL of PBS buffer, then centrifuged 15 min at 4000×g. Wet weight cells were stored at −20 °C.

#### 2.4.2. Purification

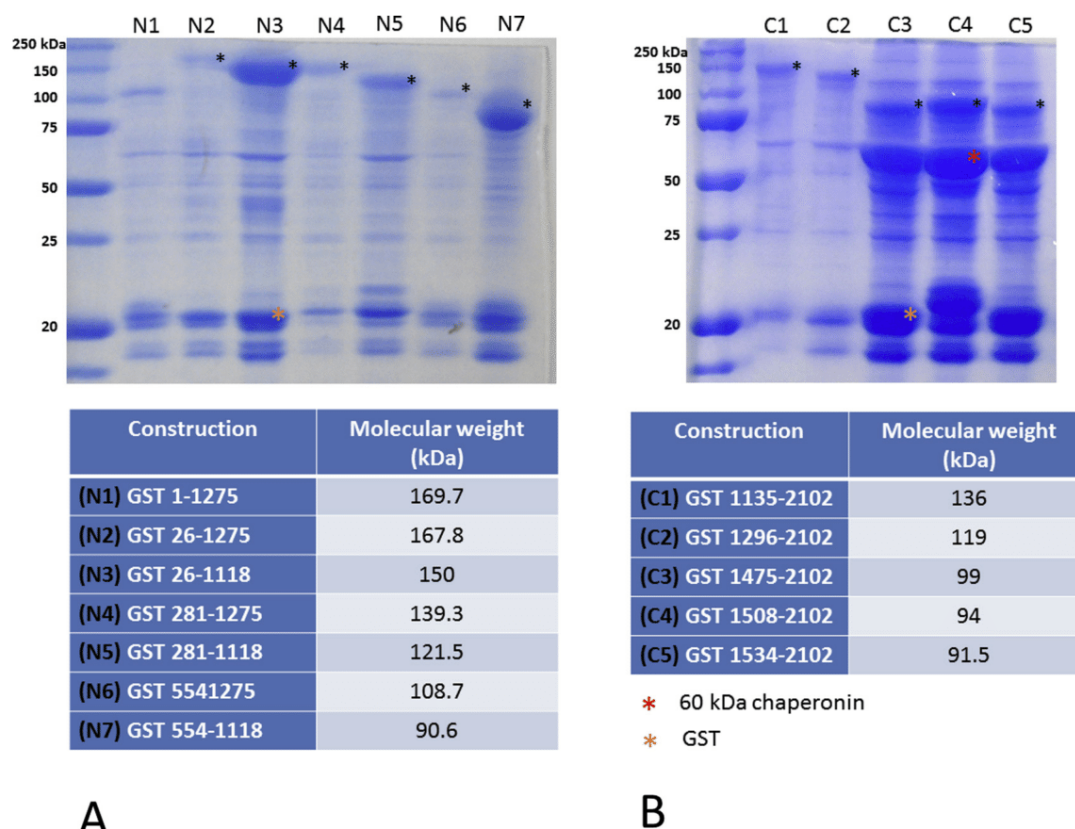
The entire process was carried out at 4 °C. Pellets from a 5 L culture were resuspended in 200 mL of lysis buffer composed of 25 mM Tris–HCl pH 8, 500 mM NaCl, 10% glycerol, 1 mM EDTA, 2 mM DTT, 2 tablets of protease inhibitor cocktail (ref. 11836170001, Roche) and about 8000 U of the DNase Benzodase (ref. 70746-3, Merck-Millipore). Resuspended pellets were incubated with 1 mg/mL of lysozyme for 1 h. Lysis was then completed with sonication about 3 min with pulses of 3 s at 40% amplitude. The lysate was clarified by centrifugation for 30 min at 40,000×g, filtered with a 0.45 μm PES membrane filter and incubated with 1 mL of Protino Glutathione Agarose 4B beads (Macherey Nagel) under agitation overnight. The resin was packed in a column and connected to an AKTA purifier 10 liquid chromatography system. The beads were washed at 1 mL/min until OD<sub>280nm</sub> baseline was reached (25 mM Tris–HCl pH 8, 500 mM NaCl, 10% glycerol, 1 mM EDTA, 2 mM DTT). Proteins were eluted with wash buffer plus 50 mM glutathione at 0.3 mL/min and collected in fractions

of 1 mL. Fractions containing purified protein were pooled, concentrated by ultrafiltration to a volume of 1 mL and incubated with 40 μg of GST-tagged PreScission protease for 2.5 h at room temperature. The sample was loaded on a size-exclusion chromatography (SEC) column (Superose 6 10/300 GL Increase, GE Healthcare), pre-equilibrated in 25 mM Tris–HCl pH 8, 500 mM NaCl, 10% glycerol, 1 mM EDTA, 2 mM DTT. The fractions containing eluted proteins were pooled and GST-tagged proteins were finally removed with a GST-Trap FF 1 mL column (GE Healthcare). The purified protein in the flow-through was concentrated by ultrafiltration to a volume of 500 μL. Each purification step was analyzed by 14% polyacrylamide SDS–PAGE.

### 3. Results and discussion

#### 3.1. Bioinformatics prediction of constructs amenable to *E. coli* expression

Full-length PI4KA can be produced in the baculovirus system and solubilized and purified with detergent in a catalytically active form [17]. N-terminally truncated constructs produced in *E. coli* have been sporadically reported also to display catalytic activity [18]. However, no clear domain boundaries could be assigned until our and our collaborators' recent work mapping regions of PI4KA functionally necessary to HCV replication [14]. A remarkable finding of that work was that the whole N-terminal 1100-residue region of PI4KA is dispensable for supporting HCV replication. In



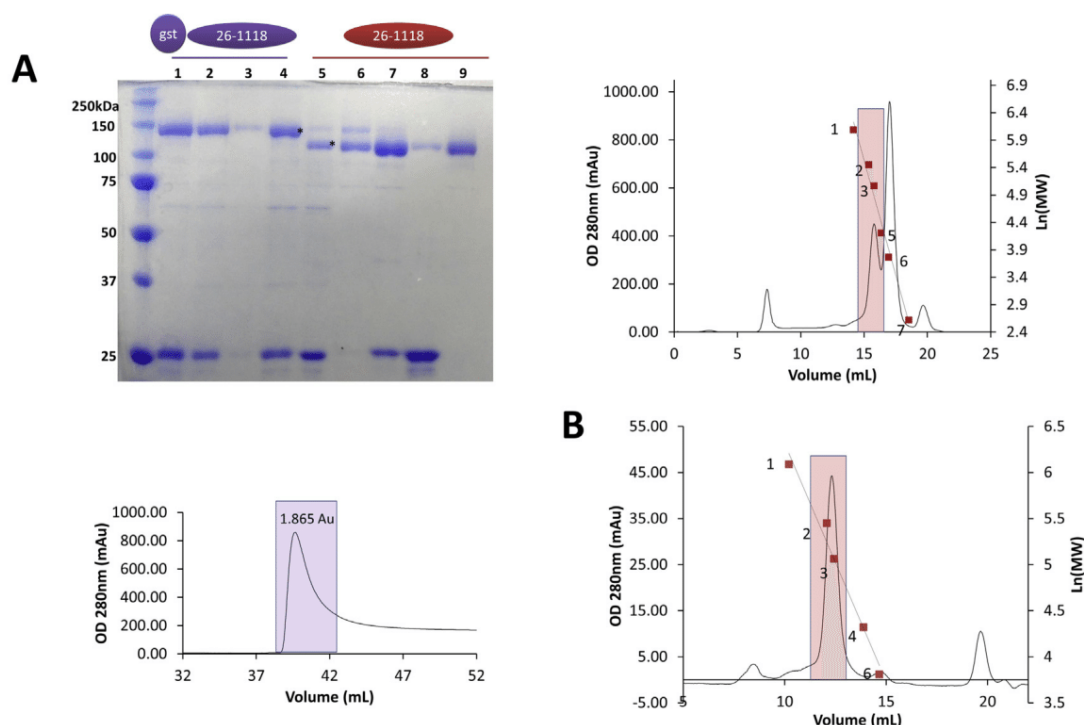
**Fig. 2.** Pull-down of soluble GST-tagged PI4KA constructs after small-scale expression in *E. coli* BL21(DE3)C41 analyzed by SDS–PAGE. Glutathione beads were deposited after incubation with the soluble fraction of a lysed 30-mL culture. Stars indicate proteins that were further confirmed by mass spectrometry. (A) Constructs from the N-terminal regulatory part of PI4KA. The DI domain (1135–1275) seems to abrogate the soluble production of the constructs regardless of the important molecular weight of the constructs. (B) Constructs from the C-terminal part including regulatory and catalytic domains. The overproduction of chaperonin (red star) indicates that the three smaller constructs are likely not folded. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the course of that work we also revised the seminal work of Gehrmann et al. [19] and proposed a new annotation of residues 1130–2102. In particular, we suggested that a subdomain (which we dubbed DI) may encompass residues 1130–1280 based on secondary structure and disorder predictions. Using similar methods, we note here (Fig. 1A) three segments in the 1100 first residues of PI4KA with predicted disorder and low predicted secondary structure and/or low consensus on secondary structure prediction: The *ca* 30 first residues, a segment comprising residues 200–320, and a short stretch around residue 550. We therefore designed as putative domains or sets of domains capable of autonomous folding a series of constructs for expression in *E. coli* (Fig. 1B). First, N-terminal PI4KA constructs starting at residue 1, 26, 281 or 554 and ending either at residue 1118 (just upstream of DI) or at residue 1275 (i.e., adding DI). Second, C-terminal constructs corresponding to truncations either competent for supporting HCV replication (1136–2102) [14] or devoid of DI.

### 3.2. Small-scale expression tests

We cloned these constructs as GST-tagged fusions for expression in *E. coli* strain BL21(DE3)C41 at 18 °C. This choice was dictated by our previous extensive screening of expression conditions with other PI4KA constructs showing that a combination of these tag and strain and an expression temperature of 12–18 °C were most favorable (Taveneau & Bressanelli, unpublished). However, as in our previous work soluble expression was too low in all cases to be detected in cell lysis supernatants. Still in our experience it is possible to perform biochemical and even

structural work on recombinant, affinity-tagged proteins in these conditions, provided that (i) expression is enough for the proteins to show up after affinity bead pulldown as sizable bands by Coomassie staining. (ii) The proteins are well behaved in solution in subsequent steps. We therefore screened the best expression constructs by GST pulldown of soluble extracts, choosing a buffer for cell lysis and pulldown that would minimize the risks of solubilizing a GST-tagged misfolded construct, e.g. a buffer without detergent (see Section 2). For N-terminal constructs, removing the first putatively disordered segment improves soluble expression noticeably (Fig. 2A, compare N2 with N1). The most striking observation is that N-terminal constructs including DI display far less soluble expression than equivalent constructs without DI (Fig. 2A, compare N3 with N2, N5 with N4 and N7 with N6). In contrast, soluble expression of C-terminal constructs is not improved by removal of DI (Fig. 2B, compare lane C2 with C1). However, expression is low even with DI (C1). Expression is improved by further deletions in the C-terminal constructs (C3–C5) but there is a simultaneous and higher boost in the expression of lower molecular weight proteins, the most prominent of which is identified by MALDI-TOF mass spectrometry as *E. coli* chaperonin (Fig. 2B). In our hands PI4KA constructs displaying this expression pattern are hard to purify from chaperones, must be kept soluble with detergents and mostly elute in the void volume in size exclusion chromatography. This behavior suggests that such constructs are not properly folded and/or display large hydrophobic patches and/or require a detergent micelle for solubility and/or are prone to massive aggregation. These results therefore suggest that DI is necessary for proper folding of the C-terminus of PI4KA and is thus



**Fig. 3.** Large-scale purification of PI4KA construct 26–1118. (A) Progress of the purification of 26–1118, step by step, analyzed on 14% SDS–PAGE. Lanes are designated as follows: Lane 1–3, elution of GST\_26–1118 purified by glutathione affinity chromatography. The chromatogram of the protein elution is displayed below the gel and was recorded at a wavelength of 280 nm. The integrated OD of the peak is indicated; lane 4, pooled fraction from the glutathione affinity chromatography; lane 5, pooled fraction from glutathione affinity chromatography after concentration and digestion by the PreScission protease. Loaded sample volume was adjusted so that the same proportion of pooled fraction was deposited as on lane 4; lane 6–8, eluted fractions of the Superose 6 10/300 GL Increase column size exclusion chromatography containing 26–1118. The corresponding chromatogram is displayed on the right of the gel, and was recorded at a wavelength of 280 nm. Elution volumes and Ln(MW) of protein standards are indicated by red squares on this chromatogram and that of (B). 1 – ferritin (440 kDa), 2 – catalase (232 kDa), 3 – aldolase (158 kDa), 4 – conalbumin (75 kDa), 5 – albumin (67 kDa), 6 – ovalbumin (43 kDa), 7 – ribonuclease (13.7 kDa); lane 9, final sample after removal of GST-tagged material by glutathione affinity chromatography and concentration to 0.5 mg/mL. (B) SEC analysis of purified 26–1118 on a Superdex 200 10/300 GL Increase column. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



structurally, as it is functionally [14], a part of the C-terminus of PI4KA. Furthermore, among the N-terminal constructs without DI soluble expression is much higher cutting after the first or in the third putatively disordered regions than in the second (Fig. 2A, compare N3 and N7 to N5).

### 3.3. Large-scale production and purification of construct 26–1118

In view of this, we further worked at producing and purifying construct 26–1118, keeping all buffers without detergent. After lysis and lysate clarification, the GST-tagged 26–1118 was not readily bound to and eluted from glutathione agarose beads. In particular most of it could not be eluted from the beads in our initial buffer conditions. We therefore screened a number of buffers to finally settle on 25 mM Tris–HCl pH 8, 500 mM NaCl, 10% glycerol, 1 mM EDTA, 2 mM DTT, the two most important parameters being DTT to obtain quantitative binding and high NaCl to obtain quantitative elution. To obtain enough starting material, we switched from flask cultures to 5-liter fermentor cultures and from LB to the richer 2YT medium. With the protocol detailed in the methods section, this yields about 37 g of wet weight cells per 5-liter 2YT fermentor. After lysis, lysate clarification and binding to glutathione agarose beads in our optimized buffer conditions, the GST-tagged 26–1118 is the major elution product together with GST (Fig. 3A, bottom, and top left, lanes 1–3). Whether the high quantity of GST is due to abortive translation or to spontaneous cleavage of the linker between GST and 26–1118 is not known. At this stage we can estimate for the first time the amount of 26–1118 that can be recovered in these production conditions. The total OD<sub>280</sub> integrated from the elution peak is 1.865 (Fig. 3A, bottom) and gel densitometry indicates that about 50% of this is GST<sub>26–1118</sub>. Given the respective absorption coefficients of GST and GST<sub>26–1118</sub> (computed from their sequences), this yields a total amount of 0.72 mg of GST<sub>26–1118</sub>, of which 0.59 mg is 26–1118 and 0.13 mg is the GST tag. Despite this low level of soluble expression, the recovered sample proved to be of good quality: The pooled GST<sub>26–1118</sub> (Fig. 3A, top left, lane 4) is readily concentrated and GST cleaved off with GST-tagged PreScission protease (Fig. 3A, top left, lane 5). Subsequent size exclusion chromatography shows that 26–1118 quantitatively elutes as a single peak well after the Superose 6 column's void volume (Fig. 3A, right, and top left, lanes 6–8). However, a consequence of this elution volume and of the dimeric state of GST is that 26–1118 is incompletely separated from GST with our Superose 6 10/300 GL Increase column. GST and GST-tagged PreScission are further removed by glutathione affinity chromatography, yielding pure 26–1118 that can be concentrated to 0.2–5 mg/mL (Fig. 3A, lane 9). The final yield of the purification shown on Fig. 3A is of 0.25 mg of pure 26–1118 starting from a 5-liter 2YT fermentor. Hence we recover a little over 40% of the initially pulled down 26–1118. When checked against globular protein standards, 26–1118 (123 kDa) elutes in SEC after catalase (232 kDa) and right before aldolase (158 kDa). This is observed with the purification chromatography medium and buffer (Fig. 3A, right) and also with pure 26–1118 and a Superdex 200 10/300 GL Increase column either in the purification buffer (Fig. 3B) or with the purification buffer brought to 1 M NaCl (not shown), suggesting that there are no residual interactions with chromatography media. Thus 26–1118 consistently elutes in SEC as a compact, soluble protein, indicative of proper folding, and is furthermore monomeric and significantly non-globular. A fluorescence-based thermal shift assay shows that 26–1118 displays a single thermal denaturation transition with a temperature of half denaturation of 45 °C (not shown).

## 4. Conclusions

Although enzymes of the PI4KIII family must associate with the cytosolic leaflet of an intracellular membrane to perform their lipid kinase activity, they are synthesized in an inactive conformation that does not spontaneously associate with membranes [1]. They can thus be called conditional peripheral membrane proteins. From a functional viewpoint this makes it hard to distinguish between recruitment to membranes and activation. When producing fragments for biochemical and biophysics studies, it is also difficult to distinguish overproduced constructs that are insoluble because of misfolding from those that are poorly soluble because they represent a membrane-associating conformation of the protein. Here we show that a PI4KA fragment comprising almost 1100 residues of the protein's N-terminus can be produced in *E. coli* and purified in a fully soluble form, suggesting that it is properly folded. This fragment elutes as a single monomeric peak by SEC and is clearly amenable to structural characterization despite the low quantity of available purified material. Indeed, further biophysical characterization to identify conditions promoting stability and robotics crystallization trials requiring very little material have been undertaken. To our knowledge, this is the first time that such a well-behaved PI4KA fragment is successfully purified. Remarkably, this 26–1118 fragment is almost complementary to a 1101–2102 fragment that was shown to be catalytically functional in a cellular context [14]. This suggests that PI4KA can be described as organized in two modules: The C-terminal one is an activation and catalysis module, while the N-terminal one is a regulatory module that has been implicated in recruitment to the plasma membrane where PI4KA-synthesized PI4P is needed [12]. This N-terminal module may be further subdivided, as suggested by the overproduction of soluble fragment 554–1118. In conclusion, our work describes the first production of a biochemically well behaved fragment of PI4KA and will allow further structural study of its N-terminal regulatory module, an element of central importance in this tightly regulated enzyme.

## Acknowledgements

This work was supported by ANRS (France REcherche Nord & sud SIDA – hiv Hépatites: FRENSh) through a grant to SB and a pre-doctoral fellowship to CT and by the French Infrastructure for Integrated Structural Biology (FRISBI) [ANR-10-INSB-05-01]. We thank Volker Lohmann for the kind gift of the pFastBacDual plasmid encoding full-length PI4KA, Nathalie Scrima and Stéphane Roche for help in early stages of this work, and Sonia Fieulaine for assistance with gel densitometry. We acknowledge the Structural and Functional Proteomics platform of the Jacques Monod Institute (CNRS and Paris-Diderot University, Paris, France) for MALDI-TOF analyses and Eric Jacquet and Naïma Nhiri (CNRS, ICSN, Gif-sur-Yvette, France) for the fluorescence-based thermal shift assay.

## References

- [1] T. Balla, Phosphoinositides: tiny lipids with giant impact on cell regulation, *Physiol. Rev.* 93 (2013) 1019–1137, <http://dx.doi.org/10.1152/physrev.00028.2012>.
- [2] A. Balla, T. Balla, Phosphatidylinositol 4-kinases: old enzymes with emerging functions, *Trends Cell Biol.* 16 (2006) 351–361, <http://dx.doi.org/10.1016/j.tcb.2006.05.003>.
- [3] J.R. Brown, K.R. Auger, Phylogenomics of phosphoinositide lipid kinases: perspectives on the evolution of second messenger signaling and drug discovery, *BMC Evol. Biol.* 11 (2011) 4, <http://dx.doi.org/10.1186/1471-2148-11-4>.
- [4] D.A. Fruman, R.E. Meyers, L.C. Cantley, Phosphoinositide kinases, *Annu. Rev. Biochem.* 67 (1998) 481–507, <http://dx.doi.org/10.1146/annurev.biochem.67.1.481>.

- [5] N.-Y. Hsu, O. Ilnytska, G. Belov, M. Santiana, Y.-H. Chen, P.M. Takvorian, C. Pau, H. van der Schaar, N. Kaushik-Basu, T. Balla, C.E. Cameron, E. Ehrenfeld, F.J.M. van Kuppeveld, N. Altan-Bonnet, Viral reorganization of the secretory pathway generates distinct organelles for RNA replication, *Cell* 141 (2010) 799–811, <http://dx.doi.org/10.1016/j.cell.2010.03.050>.
- [6] S. Reiss, I. Rebhan, P. Backes, I. Romero-Brey, H. Erfle, P. Matula, L. Kaderali, M. Poenisch, H. Blankenburg, M.-S. Hiet, T. Longerich, S. Diehl, F. Ramirez, T. Balla, K. Rohr, A. Kaul, S. Bühler, R. Pepperkok, T. Lengauer, M. Albrecht, R. Eils, P. Schirmacher, V. Lohmann, R. Bartenschlager, Recruitment and activation of a lipid kinase by hepatitis C virus NS5A is essential for integrity of the membranous replication compartment, *Cell Host Microbe* 9 (2011) 32–45, <http://dx.doi.org/10.1016/j.chom.2010.12.002>.
- [7] N. Altan-Bonnet, T. Balla, Phosphatidylinositol 4-kinases: hostages harnessed to build panviral replication platforms, *Trends Biochem. Sci.* 37 (2012) 293–302, <http://dx.doi.org/10.1016/j.tibs.2012.03.004>.
- [8] S. Reiss, C. Harak, I. Romero-Brey, D. Radujkovic, R. Klein, A. Ruggieri, I. Rebhan, R. Bartenschlager, V. Lohmann, The lipid kinase phosphatidylinositol-4 kinase III alpha regulates the phosphorylation status of hepatitis C virus NS5A, *PLoS Pathog.* 9 (2013) e1003359, <http://dx.doi.org/10.1371/journal.ppat.1003359>.
- [9] J.U. Flanagan, P.R. Shepherd, Structure, function and inhibition of the phosphoinositide 3-kinase p110 $\alpha$  enzyme, *Biochem. Soc. Trans.* 42 (2014) 120–124, <http://dx.doi.org/10.1042/BST20130255>.
- [10] J.E. Burke, A.J. Inglis, O. Perisic, G.R. Masson, S.H. McLaughlin, F. Rutaganira, K.M. Shokat, R.L. Williams, Structures of PI4KIII $\beta$  complexes show simultaneous recruitment of Rab11 and its effectors, *Science* 344 (2014) 1035–1038, <http://dx.doi.org/10.1126/science.1253397>.
- [11] J.E. Burke, O. Perisic, R.L. Williams, Allosteric activation of PI3K $\alpha$  by oncogenic mutations, *Oncotarget* 4 (2013) 180–181.
- [12] F. Nakatsu, J.M. Baskin, J. Chung, L.B. Tanner, G. Shui, S.Y. Lee, M. Pirruccello, M. Hao, N.T. Ingolia, M.R. Wenk, P. De Camilli, PtdIns4P synthesis by PI4KIII $\alpha$  at the plasma membrane and its impact on plasma membrane identity, *J. Cell Biol.* 199 (2012) 1003–1016, <http://dx.doi.org/10.1083/jcb.201206095>.
- [13] Y.-S. Lim, S.B. Hwang, Hepatitis C virus NS5A protein interacts with phosphatidylinositol 4-kinase type III alpha and regulates viral propagation, *J. Biol. Chem.* 286 (2011) 11290–11298, <http://dx.doi.org/10.1074/jbc.M110.194472>.
- [14] C. Harak, D. Radujkovic, C. Taveneau, S. Reiss, R. Klein, S. Bressanelli, V. Lohmann, Mapping of functional domains of the lipid kinase phosphatidylinositol 4-kinase type III alpha involved in enzymatic activity and hepatitis C virus replication, *J. Virol.* 88 (2014) 9909–9926, <http://dx.doi.org/10.1128/JVI.01063-14>.
- [15] O. Vadas, J.E. Burke, X. Zhang, A. Berndt, R.L. Williams, Structural basis for activation and inhibition of class I phosphoinositide 3-kinases, *Sci. Signal.* 4 (2011), <http://dx.doi.org/10.1126/scisignal.2002165>, re2.
- [16] A. Biegert, C. Mayer, M. Remmert, J. Söding, A.N. Lupas, The MPI Bioinformatics Toolkit for protein sequence analysis, *Nucleic Acids Res.* 34 (2006), <http://dx.doi.org/10.1093/nar/gkl217>, W335–339.
- [17] J. Borawski, P. Troke, X. Puyang, V. Gibaja, S. Zhao, C. Mickanin, J. Leighton-Davies, C.J. Wilson, V. Myer, I. Cornellataracido, J. Baryza, J. Tallarico, G. Joberty, M. Bantscheff, M. Schirle, T. Bouwmeester, J.E. Mathy, K. Lin, T. Compton, M. Labow, B. Wiedmann, L.A. Gaither, Class III phosphatidylinositol 4-kinase alpha and beta are novel host factor regulators of hepatitis C virus replication, *J. Virol.* 83 (2009) 10058–10074, <http://dx.doi.org/10.1128/JVI.02418-08>.
- [18] Z.A. Knight, B. Gonzalez, M.E. Feldman, E.R. Zunder, D.D. Goldenberg, O. Williams, R. Loewith, D. Stokoe, A. Balla, B. Toth, T. Balla, W.A. Weiss, R.L. Williams, K.M. Shokat, A pharmacological map of the PI3-K family defines a role for p110alpha in insulin signaling, *Cell* 125 (2006) 733–747, <http://dx.doi.org/10.1016/j.cell.2006.03.035>.
- [19] T. Gehrman, H. Gülkan, S. Suer, F.W. Herberg, A. Balla, G. Vereb, G.W. Mayr, L.M. Heilmeyer Jr, Functional expression and characterisation of a new human phosphatidylinositol 4-kinase PI4K230, *Biochim. Biophys. Acta* 1437 (1999) 341–356.

## 2) Complément d'article : Expression et solubilité des constructions C-terminales

La solubilité et le comportement en solution des différentes constructions présentées dans l'article ont été finement évalués par le biais de la mise au point de leur expression et de leur purification. Elles ont été étudiées en fusion avec une étiquette GST mais aussi avec une étiquette hexa-histidine en N-terminal. Dans un premier temps, les protéines de fusion ont été purifiées en batch afin de déterminer les conditions favorables à leur stabilité, en particulier les conditions d'expression et la nécessité de détergent. Par la suite, ces tests ont été adaptés à des purifications à grandes échelles.

Nous nous sommes plus particulièrement attardés sur les constructions C-terminales, plus difficiles à exploiter que les constructions N-terminales qui nous ont permis très rapidement de cibler une construction unique : 26-1118. En effet, seuls des résultats préliminaires tardifs nous ont permis de comprendre que seules des constructions DI-2102 ou DII-2102 étaient certainement exploitables. J'ai ainsi testé de nombreuses conditions de purification en ce qui concerne les constructions C-terminales plus petites : présence de cofacteurs tels que l'ATP et le magnésium, présence de détergent (Triton X-100, LDAO, lauryl sarcosyl) et mutants sans l'hélice C-terminale (notés dc17). Les conditions d'expression et de purification ainsi que les résultats obtenus pour les constructions C-terminales sont détaillés dans les annexes 4 et 5 respectivement ; pour les constructions N-terminales en annexes 6 et 7.

### **i) Matériel et méthode**

#### **▪ Mise au point de l'expression des protéines en *E.Coli***

Un grand nombre de souches modifiées de *E.Coli* a été testé pour la mise au point de l'expression de chacune des constructions :

- **BL21(DE3)<sup>116</sup>** : souche classique d'*E.Coli* modifiée utilisée pour l'expression. Les souches suivantes sont dérivées de BL21(DE3). L'expression des protéines d'intérêt est inductible par ajout d'IPTG. En effet, la synthèse de polymérase est médiée par un promoteur Lac.
- **C41(DE3)<sup>117</sup>** : souche supportant mieux la surexpression de protéines toxiques et membranaires. En effet, elle possède des mutations qui préviennent de la mort cellulaire liée à la présence de protéines toxiques
- **Star(DE3)<sup>118</sup>** : une mutation dans la RNase E permet une dégradation plus lente des ARNs
- **Rosetta(DE3)<sup>119</sup>** : souche améliorée afin de palier à la présence de codons rare (AUA, AGG, AGA, CUA, CCC, GGA). Cette souche nécessite une sélection au chloramphénicol (0.34µg/mL) afin de ne conserver que les bactéries contenant les vecteurs codant pour les différents ARNt.

- **ArcticExpress (DE3)**: bactérie modifiée afin d'augmenter la solubilité des protéines surexprimées en augmentant la quantité de protéine synthétisées à basse température (12°C)
- **pGro7/BL21**<sup>120</sup>: la souche de *E.Coli* est implémenté en plasmides codant pour les chaperonnes GroES et GroEL. Les gènes codant pour ces chaperonnes sont sous le contrôle du promoteur araB inducible par ajout de L-Arabinose (200µg/mL) et le vecteur les portant sous résistant au chloramphénicol (0.34µg/mL).

Chacune des protéines de fusion a été exprimée dans les mêmes conditions que celles de l'article. Seule le volume de culture, la température d'induction et la concentration en IPTG utilisés sont variables (voir annexe 4).

#### ▪ Tests de purification en batch

Le culot bactérien est repris dans 100 µL de tampon de lyse par unité d'absorbance à 600nm. Les bactéries resuspendues sont ensuite lysées par sonication avec 8 impulsions de 2 secondes d'amplitude 40. Après centrifugation à 20000xg pendant 20 minutes, le surnageant est incubé avec 20µL de résine glutathione sepharose 4B GE healthcare pour les protéines de fusion GST ; 20µL de résine Ni-NTA pour les protéines de fusions hexa-histidines. La résine est ensuite incubée pendant 2x10 minutes dans 600µL de tampon de lavage. La protéine est finalement éluée dans 50µL de tampon d'éluion (300 mM imidazole ou 20 mM glutathion).

Tous les tests de purification des constructions N-terminales ont été réalisés dans 25 mM Tris-HCl pH7.4, 100mM NaCl, 5% glycérol. En ce qui concerne les constructions C-terminales, les tampons utilisés ont varié. Leurs compositions sont détaillées en annexe 5.

#### ▪ Test de purification à grande échelle (constructions GST seulement)

Les constructions C-terminales ont donné beaucoup de faux espoirs après les tests en batch. C'est pourquoi un grand nombre de conditions ont été testées à grande échelle. Afin de simplifier la présentation des résultats, seuls les plus significatifs sont exposés dans ce chapitre. Ainsi, je me suis tout particulièrement focalisée sur la présence de détergent et de cofacteurs (voir annexe 5). Les tests à grande échelle sont toujours réalisés suivant le protocole détaillé ci-dessous, excepté en ce qui concerne la construction 26-1118. En effet la mise au point de la purification de ce fragment sera détaillée dans le chapitre suivant (Partie I.A.2).



Les cultures à grande échelle ont été réalisées dans des volumes variant de 0,5 à 2L par erlenmeyer avec ergots (de 3L pour les volumes de 0,5 à 1L ; de 5L pour les volumes de 2L). Après centrifugation de 3500g pendant 20 minutes, les culots bactériens sont repris dans 20mL de tampon de lyse par litre de culture. La lyse est réalisée par sonication avec des pulses de 3 secondes d'amplitude 30 et ce pendant 2x1minute puis 2x1 minute. 800 unités de benzonase ainsi qu'une tablette de cocktail d'anti-protéases ont été ajoutées au préalable par 20 mL de bactéries resuspendues. Le lysat est ensuite centrifugé pendant 30 minutes à 40000g puis le surnageant est filtré sur membrane PES 0,45µM. Le surnageant est ensuite chargé à 0,5 mL.min<sup>-1</sup> sur colonne pré-packée GST-trap FF au préalable branchée à une HPLC Akta purifier 10 réfrigérée à 4°C. Le lavage est ensuite effectué à 1mL.min<sup>-1</sup> dans le tampon approprié puis l'élution à 0,5mL.min<sup>-1</sup> avec 30mM glutathion tamponné à pH8. La protéine est collectée dans des fractions de 1mL.

Suivant le comportement de la protéine après l'étape de chromatographie d'affinité, certains des tests suivant ont été effectués afin d'évaluer le bon repliement de la protéine :

- Des tests de clivage de l'étiquette GST à la PreScission.

20µg de protéase sont ajoutés aux fractions contenant de la protéine d'intérêt au préalable regroupées. Un aliquot est prélevé après 1heure, 1heure30, 2heures et 3 heures d'incubation à température ambiante et après une nuit à 4°C puis déposé sur SDS-PAGE.

- Injection sur gel filtration

Afin d'évaluer le bon repliement de certaines des constructions avec ou sans étiquette GST, la protéine de fusion a été injectée sur gel filtration avec un flux de 0.25 à 0.3 mL.min<sup>-1</sup>. Deux types de chromatographie d'exclusion ont été utilisées : Superdex 200 10/300 et une Superose 6 10/300. Dans le cas des constructions avec étiquette GST, la dimérisation de l'étiquette a été prise en compte dans le choix du tamis moléculaire à utiliser.

- Tests de concentration

Les constructions C-terminales se comportant assez mal, il a été réalisé dans de rares cas des tests de concentration. Ils ont été effectués dans des concentrateurs Amicon avec une membrane de cellulose. La taille des pores a été choisie en fonction de la taille de la protéine et de la micelle de détergent si besoin.

Chacune des étapes de purification a été analysée sur SDS-PAGE.

## ii) Résultats

### ▪ Constructions C-terminales

J'ai entrepris d'exposer ici les résultats les plus significatifs quant aux constructions C-terminales. Les résultats concernant les constructions 1135/1293-2102 sont seulement préliminaires. MW et Rs correspondent respectivement au poids moléculaire et au rayon de stokes calculés à partir du volume de rétention des protéines sur la chromatographie d'exclusion.

	Constructions	Meilleur expression soluble	Solubilité avec Tag		Solubilité sans tag	
			Sans détergent	Avec détergent	Sans détergent	Avec détergent
	GST 1534-2102	Rosetta 18°C C41 12°C	Chaperonne	Non testé	Non testé	Non testé
	GST 1508-2102	Rosetta 18°C C41 12°C	Chaperonne	Non testé	Volume mort	Co-elution DNaK
	GST 1472-2102	Rosetta 18°C C41 12°C	Chaperonne	Non testé	Volume mort	MW = 289kDa Rs = 5.7nm 0,5% triton X-100
<b>Résultats préliminaires</b>	6his d1292	Rosetta 12°C	✓	Non testé	Non testé	Non testé
	GST d1135	pGro7 12°C	✓	Non testé	Non testé	Non testé

Au tout début de ma thèse, nous nous sommes focalisés sur des constructions ARM-CAT, soit les constructions 1472/1508/1534-2102. En effet, les PI3Ks et PI4K $\beta$  de structure connue sont solubles sous cette forme, sans détergent et sans leurs sous-unités régulatrices. Cependant, ces constructions ne sont pas solubles, que ce soit en détergent ou non : élution partielle lors des tests en batch, clivage partiel de l'étiquette, mauvaise solubilité sans l'étiquette GST, présence massive de chaperonne, grande perte de protéine à la concentration, volume de rétention proche du volume mort lors du passage sur chromatographie d'exclusion. La présence de cofacteur (ATP, magnésium) ainsi que des constructions sans l'hélice C-terminale, que nous soupçonnons d'interagir avec les membranes, ne modifie pas la solubilité de ces constructions.

De plus, la construction 1508-2102 a aussi été exprimée

- en système *cell free* constitué d'extraits de levures et en présence de différents détergents. La construction possédait en N-terminal une étiquette hexa-histidines. Aucune expression n'a été détectée par western blot
- en cellule d'insecte SF21 par infection au baculovirus. La construction, avec une étiquette hexa-histidine en N-terminal, est bien reconnue par western blot. Cependant, toute la protéine est détectée dans le culot après lyse par sonication en présence de 1% Triton X-100

De manière surprenante, chacune de ces trois constructions était exprimée en grande quantité mais essentiellement en corps d'inclusion. Vraisemblablement, un facteur majeur entre en jeu dans la production insoluble de ces fragments : l'étiquette GST, toujours située en N-terminal de la protéine de fusion. En effet, la GST est facile à produire et augmente la solubilité des protéines. De plus, les repliements de type répétitions Armadillo sont des motifs simples à produire par une bactérie. Ainsi, si notre construction n'est pas correctement définie, il se peut que nous ayons la GST en N-terminal, suivi d'hélices mal repliées voire un début d'agrégat, suivi du reste de la protéine correctement repliée. Par conséquent, la protéine de fusion va être surexprimée et en partie être solubilisée en présence de détergent. Ces constructions, en détergent, vont ainsi très bien se comporter pendant la chromatographie d'affinité mais agréger au moment du clivage de l'étiquette. Or, nous ne savons pas comment PI4KA interagit avec les membranes : directement ou indirectement par le biais de partenaires, présence de grands patch hydrophobes ou non, passages transmembranaires, etc... Ainsi, il a été question de comprendre si ce comportement insoluble était lié à un mauvais repliement des constructions ou bien à son caractère peut être fortement hydrophobe. La présence massive de chaperonnes atteste d'un mauvais repliement de nos constructions, et explique aussi la solubilisation d'une grande quantité de ces constructions. De plus, cela explique aussi une meilleure expression de ces constructions dans la souche C41(DE3), souche supportant mieux les protéines toxiques et donc la surexpression de protéines mal repliées.

Deux constructions C-terminales plus étendues se comportent mieux en solution que les constructions ARM-CAT : 1135-2102 et 1292-2102, soit les constructions débutant avant et après le domaine putatif DI. Les résultats sont préliminaires, mais encourageants. Ces constructions se produisent en quantité moindre que celles étudiées précédemment. Cependant, nous sommes à la limite de la taille des protéines surexprimées en bactéries (136 et 119kDa respectivement). De plus, leur repliement peut être un peu plus complexe car nous sommes en présence de domaines putatifs supplémentaires adoptant certainement un repliement un peu différent que le domaine ARM. Malgré cela, ces constructions adoptent un comportement de protéine soluble:

- Pas de présence massive de chaperonne lors des tests en batch
- 1135-2102 a été produite en pGro7/BL21. Malgré la surexpression des chaperonnes GroEL/ES chez ces souches bactériennes, la protéine de fusion a été purifiée en batch sans chaperonnes
- L'étiquette GST, dans les deux cas, est clivable en petit test
- 1292-2102 est la seule construction à être produite avec une étiquette hexa-histidines en N-terminal. Les autres constructions C-terminales ne surexpriment pas dans ces conditions. Or, contrairement à l'étiquette GST, l'étiquette hexa-histidines n'augmente pas la solubilité de protéines.

Ce dernier point est d'autant plus important qu'il s'agit d'une construction sans le domaine putatif DI. Ainsi, ce dernier diminue la solubilité des constructions N-terminal, comme expliqué dans l'article précédent, mais pourraient aussi diminuer la solubilité du module C terminal. DI pourrait ainsi être un domaine autonome, peu soluble, différent du domaine putatif DII et pivot entre les modules N- et C-terminaux. Le caractère insoluble de 1135-2102 peut aussi être dû à une instabilité de DI sans cofacteurs : interactions avec une protéine régulatrice activée par l'intermédiaire de nucléotides comme il est courant chez les PI3/4Ks (Gtpases Ras, Rab), ou encore interaction directe avec les membranes. Cela expliquerait le rôle primordial de DI dans l'activité de PI4KA.

Par manque de temps, ces pistes n'ont pas pu être plus exploitées. Une expression de cette construction en cellule d'insecte ainsi que l'étude biochimique et biophysique de ces constructions pourraient apporter des données quant au mécanisme d'activité de la PI4KA.

#### ▪ Constructions N-terminales

*La mise au point de l'expression et de la purification de 26-1118 évoquée dans l'article précédent ne seront pas abordées ici, mais au cours du chapitre Partie I.B spécifique à cette protéine de fusion.*

Peu de tests supplémentaires à l'article ont été réalisés sur les constructions N-terminales. En effet, 26-1118 étant prometteuse, nous nous sommes très rapidement focalisés sur cette dernière. Seuls quelques tests de purification en batch supplémentaires avec Triton X-100 ont été réalisés sur les constructions X-1118. Globalement, leur solubilité n'a pas été beaucoup augmentée, même en ce qui concerne la construction 276-1118. Par ailleurs, la protéine de fusion 276-1118 est insoluble, cependant nous ne pouvons pas conclure quant à la présence ou pas d'un domaine N-terminal 1-300. En effet, les prédictions de structure secondaires étant difficilement interprétables, il était difficile de définir la limite N-terminale de cette construction. De plus, au vu de la construction 553-1118, il semblerait bien que le module N-terminal possède deux domaines putatifs au minimum.

Les constructions 26-1118 et 554-1118 s'expriment bien en bactérie malgré leur grande taille, 150 et 91 kDa respectivement. Cela est en accord avec nos prédictions bio-informatiques. En effet, des repliements de type répétitions HEAT sont des motifs facilement exprimés par les bactéries, à l'inverse des structures possédant des brins  $\beta$  et des repliements hélicaux plus complexes.

Avant de réaliser ces constructions, nous avons essayé d'exprimer les constructions GST-611-1088 et GST-611-1118 (voir annexe 6 et 7). Il s'avère que ces constructions ne sont pas solubles sans détergent, seule la construction 611-1088 en présence de 0.25 mM Triton X-100 présente un profil encourageant, avec un rayon de stokes expérimental de 4nm (MW théorique = 54,58kDa). Cependant, nous avons voulu éviter au maximum l'utilisation de détergent afin de faciliter les études biophysiques. Ainsi, nous n'avons pas approfondi les études de cette construction.

### 3) Mise au point de l'expression et de la purification de 26-1118 à grande échelle

#### **i) Matériel et méthode**

##### **▪ Choix de la souche bactérienne et de la température d'induction**

Dans un premier temps, la température d'induction ainsi que la souche bactérienne ont été déterminées. Au vu des résultats obtenus sur les tests en batch, nous avons choisi d'utiliser la souche C41(DE3) car tout indique que la protéine se produit en majorité sous forme de corps d'inclusion toxiques pour *E.Coli*. Afin de déterminer la température optimale de surexpression de la protéine, trois cultures de 30mL en LB-ampicilline ont été conduites sous trois conditions : induction à 21 et 23°C sur la nuit et 27°C sur 5 heures. Le protocole suivi est le même que celui de l'article : induction par ajout de 1mM d'IPTG lorsque la DO<sub>600nm</sub> atteint 0,8. Chacune des cultures a ensuite été purifiée en batch en suivant le protocole de l'article n°2. Seul le tampon de lyse est différent (25mM Tris HCl pH7,4, 500mM NaCl, 10% glycérol, 1mM EDTA). La résine a été déposée sur SDS-PAGE afin d'évaluer la quantité de protéine soluble produite.

##### **▪ Optimisation des conditions de culture : choix du milieu et culture en fermenteur**

L'expression de GST-26-1118 dans les conditions les plus simples, c'est-à-dire en erlenmeyer contenant du milieu minimum LB (Lysogeny Broth) dans un incubateur, était très fastidieuse. En effet, plus de dix litres de culture étaient nécessaires afin d'obtenir suffisamment de protéine pour sa caractérisation. Nous avons donc cherché à améliorer les conditions de culture. Pour cela, nous nous sommes adressés à Karine Blondeau, maître de conférences à l'Université Paris-Sud XI, possédant une grande expertise en culture bactérienne en fermenteur.



Afin d'obtenir une quantité plus élevée de protéine, nous avons décidé de tester différents milieux de culture en fermenteur. Ainsi, il nous était possible de surveiller et de contrôler finement les paramètres essentiels à de bonnes conditions de culture ainsi que le comportement des cultures bactériennes : pH, température, oxygénation, agitation.

Trois milieux de culture ont été testés en batch, c'est-à-dire sans enrichissement du milieu pendant la culture :

- **Lysogeny Broth-Miller**, milieu riche composé principalement d'extrait de levure et de sel
- **2YT**, milieu similaire au LB avec deux fois plus de sel et d'extraits de levure
- **Auto-inductible MAI**, milieu riche contenant trois sources de carbones différentes : glucose, lactose et glycérol. Ainsi, il existe un contrôle précis du promoteur grâce à la répression catabolique exercée par le glucose qui va être métabolisé en premier. Puis la bactérie va ensuite utiliser le glycérol comme seconde source de carbone lors de l'induction de l'expression de protéine par le lactose via sa liaison au répresseur Lac I. Ceci permet entre autre d'éviter les fuites d'expression.

Le milieu de culture M63 a quant à lui été testé en fed-batch, c'est-à-dire avec enrichissement régulier en nutriment au cours de la culture. Ce milieu est principalement enrichi en sel, en vitamine B, en minéraux et en source de carbone (glycérol).

La composition exacte de chacun de ces milieux de culture est détaillée annexe 8.

Les conditions de culture sont les suivantes :

Milieu	Préculture (37°C sur la nuit)	T°C induction	DO <sub>600nm</sub> induction	[IPTG]	Durée d'induction	Volume	agitation	pH
<b>LB</b>	Jour 1 : LB Jour2 : LB	23°C	≈ 0,8	1mM	17 heures	5L	600 rpm	Non régulé
<b>YT</b>	Jour 1 : YT Jour2 : YT		≈ 0,8	1mM	17 heures	5L		
<b>MAI</b>	Jour 1 : YT Jour2 : MAnI		≈ 0,8	/	23 heures	1L		
<b>M63</b>	Jour1 : YT Jour2 : M63		8	1mM	/	1L		

L'oxygénation a été fixée à un volume d'air par litre de culture pour un volume de milieu par minute. Si besoin, de l'anti-mousse était ajouté manuellement. Les cultures ont été finalement centrifugées de 4000 g pendant 20 minutes à 15°C, puis les culots resuspendus dans du tampon PBS 1X. Une dernière centrifugation de 15 minutes de 5000g à 4°C a été effectuée puis les culots stockés à -20°C.

Cas particulier de la culture en M63 : Le système de fed batch est mis en place au début de la seconde préculture.

## ▪ Purification

Le protocole de purification détaillé dans l'article a été très rapidement mis au point. Aucune modification majeure n'a été apportée. Cependant, si un changement dans la composition du tampon doit être mentionné, il le sera au niveau des résultats. Plusieurs étapes ont tout de même demandé des optimisations :

### - La lyse

La lyse semblait être non reproductible et difficilement transposable à de gros volumes. Nous avons ainsi testé différentes conditions de lyse en jouant essentiellement sur la solubilisation des membranes et la dislocation des acides nucléiques. Par conséquent, nous avons essayé d'optimiser notre protocole initial (lyse par sonication).

### - Fixation de la protéine à la résine de glutathion

Pendant longtemps, nous avons eu des problèmes de fixation de GST-26-1118 lors de l'étape de chromatographie par affinité. Deux incubations successives de nos lysats avec la résine de glutathion ont été nécessaires à la fixation de la majorité de la protéine d'intérêt. Par conséquent, nous avons réalisés des expériences de fixation de notre protéine dans différents tampons afin de résoudre ce problème :

<i>Tampon initial</i>	<i>Tampon 1</i>	<i>Tampon 2</i>	<i>Tampon 3</i>	<i>Tampon 4</i>
25 mM Tris pH 7,4 500 mM NaCl 10% glycérol 1mM EDTA	25 mM Tris pH 7,4 100 mM NaCl	25 mM Tris pH 7,4 100 mM NaCl 10% glycérol	25 mM Tris pH 7,4 500mM NaCl 10% glycérol	25 mM Tris pH 7,4 500mM NaCl

Pour se faire, des culots bactériens issus de 30mL de culture chacun (LB, 23°C en erlenmeyer) ont été repris dans 1mL de tampon puis soniqués pendant 1 minute à une amplitude 30, impulsion 3. Le lysat a ensuite été centrifugé à 2000xg pendant 30 minutes et le surnageant incubé avec 40µL de résine de glutathion sur la nuit à 4°C. Le lendemain matin, la résine a été lavée avec 2x1mL de tampon 25mM Tris pH7,4, 500 mM NaCl, 10% glycérol et 1mM EDTA. La résine a ensuite été déposée directement sur SDS-PAGE.

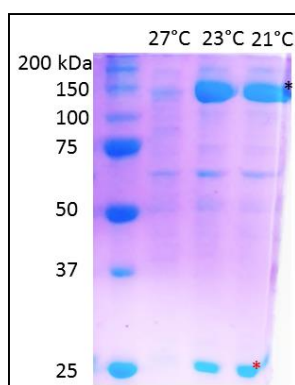
La mise au point de l'expression et de la purification de 26-1118 a été l'occasion de former Lucien Plessis, deuxième année en BTS, aux techniques de purification des protéines de fusion.

## ii) Résultats

### ▪ Expression de GST 26-1118

#### - Choix de la température

Il a été question dans un premier temps de choisir la température d'expression optimum de la protéine. Au vu de nos résultats, la protéine n'est pas soluble ou peu traduite à 27°C sur 5 heures. L'absence de GST par rapport aux tests à 21 et 23°C démontre une production très faible de la protéine de fusion. Au contraire, une expression soluble de GST 26-1118 tout à fait convenable est observée pour les expressions menées à 21 et 23°C sur la nuit.



**Figure 29- Tests de solubilité de GST 26-1118 exprimée à 21, 23 et 27°C en *E. Coli* (DE3) C41.**  
*La résine de glutathion a été déposée sur SDS-PAGE 14%.*

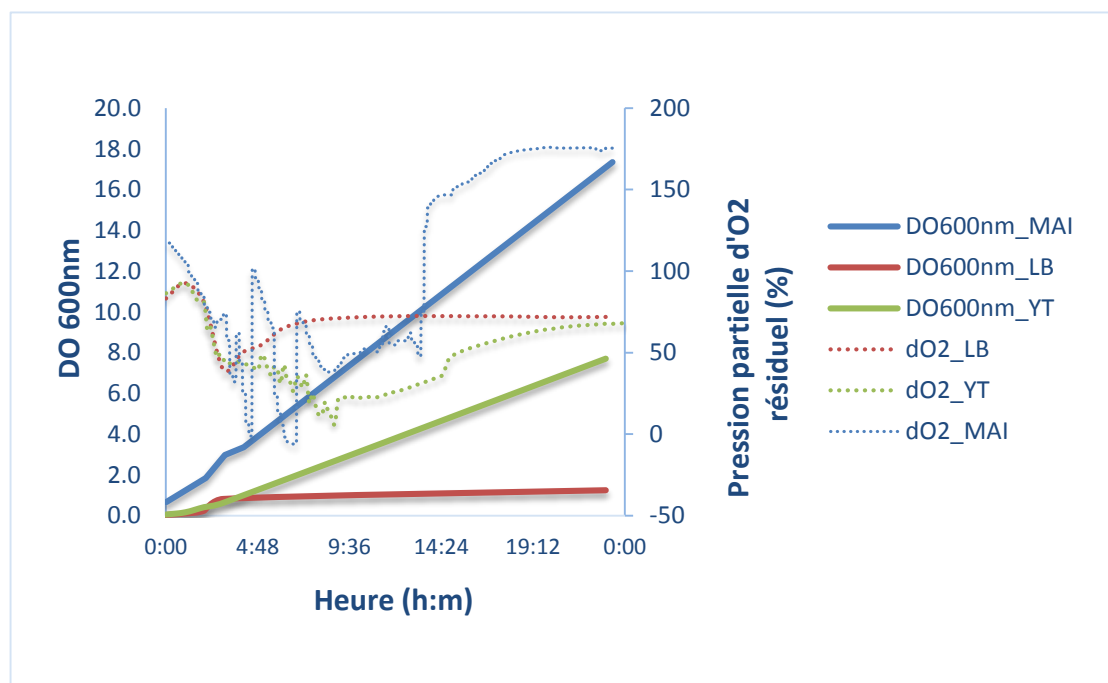
Transposer le protocole d'expression de petit à grand volume de culture n'est pas toujours réalisable. Nous avons ainsi choisi de nous placer à 23°C pour nos mises au point afin de nous placer dans des conditions d'expression optimum.

#### - Choix du milieu de culture

Au vu des résultats synthétisés dans le tableau figure 30, la souche (DE3)C41 surexprimant notre protéine nécessite un milieu riche en nutriments. En effet, en MAI, la biomasse est multipliée par 11 et la production de protéine par 3 par rapport aux cultures en milieu moins riche LB. Cela s'explique par la phase exponentielle de croissance des bactéries qui est plus longue en milieu plus riche comme nous pouvons le voir sur la figure 30 en observant la pression partielle de dioxygène résiduel, c'est-à-dire la quantité d'oxygène non consommée par la bactérie. En effet, lorsque cette pression atteint un plateau, cela signifie que le métabolisme de la bactérie est stabilisé, et par conséquent que la production de GST 26-1118 est fortement réduite.

Or, la phase de décélération précédant la phase stationnaire est atteinte au bout de 4 heures lors de culture en LB, 9 heures en YT et 17 heures en MAI. Ce plateau, ou phase stationnaire, au cours duquel la phase de croissance est nulle, est généralement détecté grâce à la densité optique à 600nm. Malheureusement, nos cultures se déroulant sur la nuit, nous n'avons pas pu réaliser de suivi affiné de la turbidité du milieu car les prélèvements sont manuels. C'est pourquoi nous n'observons pas de plateau pour la culture en milieu YT. Seules les données en MAI ont été relevées toutes les minutes. En observant la turbidité de la culture réalisée en MAI, un petit saut de densité optique vers trois heures de culture est observé. Il s'agit d'un phénomène de diauxie se produisant lorsque la bactérie a fini de consommer une de ses sources de carbone, ici le glucose.

milieu	DO <sub>600nm</sub> finale	pH final	Masse bactérie/ Litre de culture	Lyse bactérienne	Quantité protéine purifié/ Litre de culture
LB erlenmeyer	/	/	/	non	13µg/L
LB fermenteur	1.68	8.3	3grammes	non	30µg/L
YT	7.7	8.7	15 grammes	non	63µg/L
MAI	17.4	6.3	35 grammes	faible	≈ 100µg/L
M63	5.8	6.9	/	forte	Non testé



**Figure 30- Synthèse des résultats obtenus pour les tests d'expression de 26-1118 en bactérie**

**En haut.** Tableau récapitulatif des cultures réalisées en différents milieux pour l'optimisation de l'expression de la protéine d'intérêt

**En bas.** Graphique regroupant la pression partielle de dioxygène résiduel ainsi que le suivi en densité optique mesurée à 600nm en fonction du temps de culture. A noter que le profil de pression partielle d'O<sub>2</sub> en milieu MAI fluctue beaucoup de 3 à 14 heures en raison de problèmes d'agitation.

La phase exponentielle de croissance étant plus longue en MAI, une biomasse bien supérieure aux autres cultures pour une même durée d'induction est obtenue. Rapportée à la masse de bactérie utilisée, la quantité de 26-1118 après purification est plus importante dans les milieux les plus riches. Cependant, il est obtenu moins de protéine qu'attendu. Par exemple, entre milieu LB et YT, la biomasse est trois fois plus importante alors que la quantité de 26-1118 pure est seulement doublée. Compte tenu de la forte expression de corps d'inclusion et des problèmes de reproductibilité de lyse rencontrés sur des masses plus importantes de bactéries, nous pouvons difficilement évaluer la solubilité de la protéine lors de surexpression plus importante.

Les résultats obtenus pour la culture bactérienne en M63 n'ont pas été présentés car une très forte lyse des cellules a été observée dès l'induction de la surexpression de la protéine d'intérêt. Or, le taux d'oxygénation, la température et le pH étaient maintenues à des valeurs optimum pour une culture de *E.Coli* (pH neutre maintenu entre 7.2 et 6.9). Cette lyse provient donc d'un autre facteur : contamination, protéine produite avec une toxicité supérieure dans ces conditions... Malheureusement, une seule culture en M63 a pu être conduite. Un autre test devrait être réalisé afin de tirer des conclusions quant au comportement de la bactérie recombinante en milieu M63.

La construction GST-26-1118 semble adopter le comportement d'une protéine toxique ou membranaire : besoin de nutriments afin d'avoir une densité cellulaire acceptable, mort cellulaire. De plus, la lyse est observée lors de cultures en milieu à plus forte pression osmotique et à densité cellulaire élevée. Il se pourrait donc que la toxicité de la protéine provienne de la fragilisation des membranes de la bactérie. Cependant, nous ne pouvons affirmer que cette toxicité soit liée à la production de corps d'inclusion. En effet, PI4KA est une protéine périphérique aux membranes et est considérée comme n'interagissant que faiblement avec ces dernières. Cependant, quelles que soit la construction, PI4KA entière (exprimée en baculovirus) ou le domaine DI (voir partie II), le comportement est le même : il est très difficile d'isoler la protéine dans le surnageant lors de la lyse. Par conséquent, nous ne pouvons distinguer ici si 26-1118 est produite essentiellement en corps d'inclusion ou bien si elle interagit fortement avec la membrane.

Quoi qu'il en soit, *E.Coli* ne semble pas être le système d'expression le plus adapté aux différentes constructions de PI4KA étudiées. En effet, seul un système d'expression sans membrane, du type *cell free*, lèverait rapidement cette interrogation qui nous a souvent posé problème tout au long de ma thèse. Par ailleurs, ce système peut être utilisé en présence de détergent, ce qui faciliterait le criblage à effectuer si nécessaire.

Par manque de temps, la protéine a tout de même été exprimée en *E.Coli* pour réaliser les premières caractérisations du fragment.



▪ Purification de GST 26-1118

- Optimisation de la lyse

Dans toutes les conditions testées, la quantité de protéine purifiée finale reste très faible. Cependant, nous avons tout de même essayé d'extraire un maximum de protéine afin d'augmenter notre rendement. Initialement, nous lysions les bactéries résultants de petits volumes de culture ( $\approx 3$  grammes de bactéries) par sonication. Le culot était au préalable resuspendu dans un tampon riche en sel afin d'augmenter la pression osmotique et donc de faciliter la lyse mécanique. Environ 14  $\mu\text{g}$  de 26-1118 pure était obtenus (soit 44  $\mu\text{g}$  normalisé à 10g de bactéries pour faciliter la comparaison avec les autres conditions de lyse).

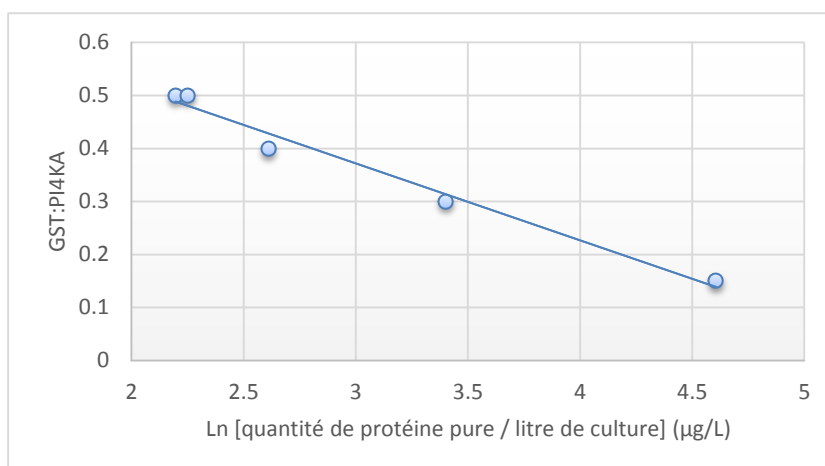
Cependant, la transposition de ce protocole à des masses beaucoup plus importantes de bactéries s'est avérée difficile. En effet, la sonication est une méthode de lyse assez violente, altérant facilement les protéines. Par conséquent, un équilibre entre sonication et méthodes de lyse plus douces afin de solubiliser les membranes entièrement et les acides nucléiques étaient indispensables à la stabilité de la protéine.

Dans le tableau ci-dessous, les avancées notables de la mise au point de la lyse des bactéries ont été répertoriées. Compte tenu de la grande quantité de bactérie lysée, l'ajout de benzonase s'est avéré essentiel à la bonne purification de la protéine. De plus, l'ajout de lysozyme incubé une heure aux cellules a significativement augmenté l'efficacité de lyse des cellules. Cependant, le rapport entre biomasse et quantité de protéine purifiée est faible.

<i>Tampon Lyse</i>	<i>Quantité bactérie</i>	<i>Volume de tampon</i>	<i>Paramètres Sonication</i>	<i>Quantité protéine pure pour 10g bactéries</i>
20mm tris pH 7.4 500 mM NaCl 5% glycérol 1 tablette anti protéase	3g	20mL	Amplitude 40 Impulsion 3 4 minutes	44 $\mu\text{g}$
+1 mM EDTA	9g	200mL	6 minutes	4,5 $\mu\text{g}$
+1mg/mL lysosyme + benzonase +2mM DTT	24g	150mL	5 minutes	12,5 $\mu\text{g}$
Incubation 1 heure avant sonication	35g	100mL	Amplitude 35 3 minutes	35,7 $\mu\text{g}$

Au cours des différentes purifications effectuées, une relation entre GST purifié et efficacité de lyse a été déterminée. En effet, plus la lyse est efficace, plus le ratio 26-118 : GST diminue lors de l'étape de chromatographie d'exclusion. Or, le fragment 26-1118 seul n'a pas été détecté dans les fractions non retenus par l'étape de chromatographie d'affinité excluant ainsi l'hypothèse de la protéolyse spontanée.

Par conséquent, cette augmentation de GST seule peut soit provenir de l'agrégation massive de la protéine observée lors de la lyse, soit d'un arrêt de traduction.



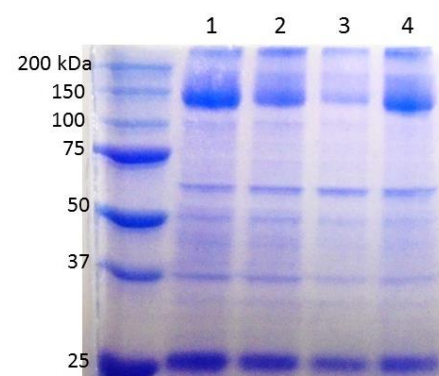
**Figure 31-** Relation observée entre la quantité de protéine purifiée et le ratio GST : 26-1118

Aucune extraction plus efficace de la protéine soluble n'a pu être mise au point lors de ma thèse. Cependant, tout semble indiquer que notre construction 26-1118 est bien un module autonome de PI4KA avec un repliement indépendant. Ainsi, le système d'expression choisi ne serait pas le système optimum. Au vu du repliement assez « simple » en motifs répétés hélicaux de ce module, il serait intéressant de tester une purification de type dénaturation/renaturation à l'urée avant de choisir un autre système d'expression.

- Fixation de la protéine de fusion à la résine de glutathion

Nous avons dans un premier temps étudié l'impact du glycérol et du sel sur la fixation de la protéine de fusion sur la résine de glutathion :

<b>Tampon 1</b>	25 mM Tris pH 7.4 100 mM NaCl	<b>Tampon 3</b>	25 mM Tris pH 7.4 500mM NaCl 10% glycérol
<b>Tampon 2</b>	25 mM Tris pH 7.4 100 mM NaCl 10% glycérol	<b>Tampon 4</b>	25 mM Tris pH 7.4 500mM NaCl

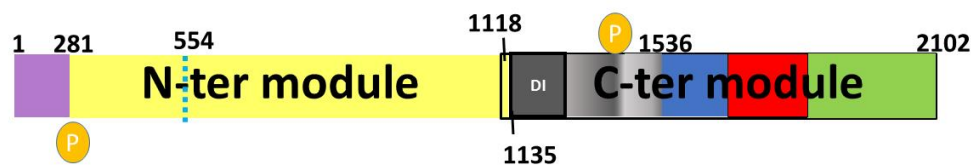


**Figure 32-** Tests de fixation de la protéine de fusion GST-25-1118 sur la résine de glutathion.

Des tests en batch similaires à ceux réalisés dans l'article n°2 ont été menés. Seuls les tampons de lyse, répertoriés dans le tableau ci-dessus, varient. La résine de glutathion a été directement déposée sur gel SDS-PAGE Tris-glycine 14%, sans étape d'élution,

Le glycérol affecte fortement la fixation de GST-26-1118 aux billes de glutathion. De plus, l'effet observé est encore plus fort quand ce glycérol est associé à du sel en concentration élevée. Or, le module 26-1118 serait composé de motif répétés de type répétition HEAT impliqué dans l'interaction avec diverses partenaires protéiques. Par conséquent, 26-1118 est probablement flexible, surtout en l'absence du module C-terminal. En augmentant la force ionique et la viscosité du tampon, le fragment pourrait donc être piégé dans une conformation masquant le site d'interaction de la GST avec le glutathion. De manière surprenante, l'effet du glycérol a pu être contré par l'ajout de 2mM de DTT. Le DTT est normalement préconisé pour une meilleure fixation de la GST à son substrat.

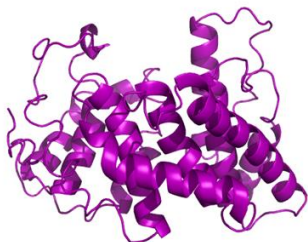
## Points importants de la partie I



DI-DIII  
Phosphorylation  
Régulation ?

Domaines ARM-Catalytique

Domaine 1-281



Riche en sérines  
phosphorylables et prolines  
Régulation ?

Module 281(?) - 1118



Répétition HEAT  
Interaction protéique ?  
Transport ?  
2 domaines ?



Interrupteur : hélice C-terminale ?  
ARM : interaction avec des partenaires ?  
Allostérie ?



# Partie II

---

Expression à grande échelle, purification et  
caractérisation biophysique de deux composantes de  
PI4KA : module N-terminal et domaine putatif DI





# I. Caractérisation du module N-terminal de PI4KA 26-1118

---

## A. Matériel et méthode

### 1) Evaluation du rayon de Stokes par chromatographie d'exclusion

La protéine 26-1118 a une forte tendance à interagir avec la matrice des colonnes de purification utilisées. Afin de caractériser la protéine d'intérêt par gel filtration, j'ai donc décidé d'injecter la protéine purifiée sur deux colonnes différentes :

<i>Tamis moléculaire</i>	<i>séparation</i>	<i>Type de résine</i>
<b>Superose 6 10/300</b>	5-5000 kDa	Agarose hautement réticulé
<b>Superdex 200 10/300 GL Increase</b>	10-600 kDa	Agarose et dextran réticulé

Les colonnes ont été équilibrées au préalable dans un tampon riche en sel afin d'éviter au plus les interactions avec les matrices : 25 mM Tris-HCl pH8, 500 mM NaCl, 10% glycérol, 2 mM DTT. L'expérience s'est déroulée sous un flux de 0,5mL.min<sup>-1</sup>.

### 2) Etude de la thermostabilité par fluorescence

*Ces études ont été menées sur la plateforme IMAGIF par Eric Jaquet et Naima Nirhi.*

L'étude de la thermostabilité des protéines par fluorescence permet de mesurer la dénaturation des protéines par un gradient de température. Pour cela, il est utilisé un colorant fluorescent, le Sypro Orange, qui interagit spécifiquement avec les résidus hydrophobes. Ainsi, lorsque la protéine est sous sa forme native, le colorant interagira peu avec les résidus hydrophobes qui sont essentiellement enfouis. Or les molécules d'eau diminuent considérablement l'émission de fluorescence par le Sypro Orange. A contrario, lors de la dénaturation thermique de la protéine, les résidus hydrophobes seront plus exposés au solvant. Ainsi, le signal de fluorescence émis par le fluorochrome sera plus fort. Par conséquent, il est facilement réalisable un criblage visant à déterminer les conditions de tampon favorable à la stabilité de la protéine en fonction de la température.

Les expériences de criblage sont réalisées sur un appareil à PCR en temps réel, le ABI 7900 HT. L'appareil permet l'excitation du Sypro Orange à 448 nm et l'enregistrement de son spectre d'émission de 500 à 600nm, mesurant ainsi les variations de fluorescence en fonction du gradient de température imposé à la protéine étudiée.

Dans un premier temps, un test de faisabilité et d'optimisation permet de définir la température de demi-dénaturation de la protéine T<sub>m</sub> ainsi que son profil de dénaturation. Pour ce premier test, il a été analysé un échantillon de 26-1118 à 0,5mg/mL suivant deux gradients de température : de 25 à 95°C puis 10 à 95°C pour préciser le profil de dénaturation de la protéine. Dans un second temps, le criblage de tampon peut être réalisé. Chaque condition est testée en duplicat sur une microplaque 96 puis, soit 48 conditions. Quatre types de tampon sont testés (Tris-HCl pH 8,5 ; Phosphate pH 7,5 ; Hepes pH 7,0 ; MES pH 6,0), avec des concentrations variables en NaCl, KCl, MgCl<sub>2</sub> et glycérol. 5 µg de protéines (tampon initial 25mM Tris-HCl pH7.4, 500 mM NaCl, 10 % glycérol, 1mM edta) ont été injectés par puits. L'échantillon initial étant trop dilué par rapport à la concentration saline à diluer (0,8mg/mL), 87mM NaCl sont à ajouter aux conditions testées initialement.

### **3) Influence du sel sur la stabilité et la solubilité de 26-1118**

Le fragment 26-1118 ayant besoin d'une forte concentration saline dans nos conditions de purification (300Mm NaCl minimum), le bruit de fond des résultats obtenus par thermo-fluorescence était assez important. Nous avons décidé de réaliser des tests de concentration dans différents tampons afin de nous assurer des résultats obtenus. Pour cela, 50µL de protéine purifiée à 0,25mg/mL (12,5µg) ont été dilués dans 500µL de tampon à tester (voir tableau ci-dessous). Le mélange est ensuite reconcentré dans 50µL finaux. Pour cela, un centricon à membrane PES de cut off 50 kDa a été utilisé avec des centrifugations successives de 14000g à 4°C.

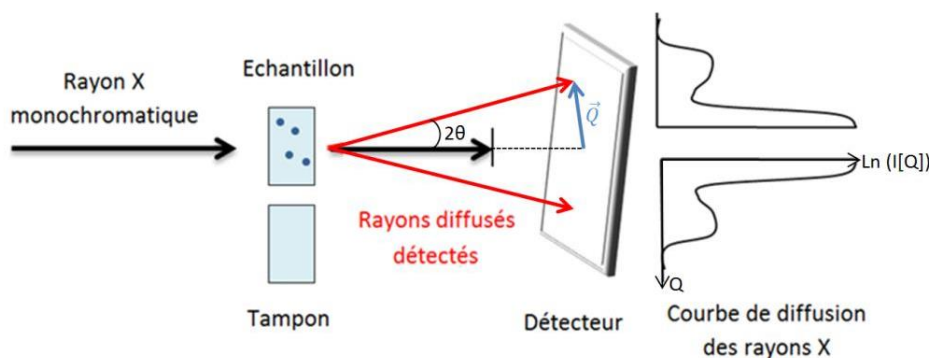
<b><i>Tampon1</i></b>	<b><i>Tampon2</i></b>	<b><i>Tampon3</i></b>	<b><i>Tampon4</i></b>	<b><i>Tampon 5</i></b>
25 mM Tris pH 7,4 500 Mm NaCl 10% glycérol	50mM NaKP pH 7,5 150 mM NaCl 2% glycérol	50mM NaKP pH 7.5 250 mM NaCl 2% glycérol	50mM NaKP pH 7.5 350 mM NaCl 2% glycérol	50mM NaKP pH 7.5 450 mM NaCl 2% glycérol

#### 4) Diffusion des rayons X aux petits angles (SAXS)

##### i) Principe<sup>122</sup>

Le SAXS est utilisé afin d'étudier les structures à basse résolution de systèmes non cristallins de différentes natures : macromolécules, polymères, systèmes inorganiques. Cette technique permet aussi de suivre des cinétiques et des changements conformationnels.

Cette méthode est basée sur la diffusion des rayons X par les électrons. Une source monochromatique de rayon X est envoyée sur l'échantillon de protéines en solution. Une grande partie de ces rayons n'interagit pas avec le dispositif expérimental ou va être diffusée dans la continuité de la source. Ils sont arrêtés par le beam stop et n'atteignent pas le détecteur. L'autre partie des rayons est quant à elle diffusée par les électrons du solvant et des protéines dans toutes les directions. Il est ainsi mesuré la somme des deux intensités par le détecteur. Par la suite, il est retranché celle des rayons diffusés par le tampon afin d'obtenir une courbe de diffusion ne dépendant que des protéines. Etant en solution, le signal est moyenné car les particules adoptent des orientations aléatoires au cours du temps et dans l'espace.



**Figure 33-** Représentation schématique d'une expérience typique de SAXS.

L'échantillon est placé dans un faisceau de rayon X. Le détecteur mesure alors l'intensité de diffusion  $I(Q)$  expérimentale. L'angle de diffusion est toujours égal à  $2\theta$ . Les courbes de diffusion obtenues sont exprimées par  $\ln[I(Q)]$  en fonction de  $Q$ , le module du vecteur de diffusion  $\vec{Q}$

Le SAXS dépend essentiellement du contraste de densité électronique entre la protéine et son solvant. En effet, l'intensité des rayons diffusés  $I(Q)$  dépend de l'amplitude de diffusion  $f(Q)$  :

$$I(\vec{Q}) = f(\vec{Q}) \cdot f^*(\vec{Q})$$

Or,  $f(Q)$  dépend de la transformée de Fourier du contraste de densité électronique. Ainsi, plus la concentration en protéine est importante, plus le contraste de densité électronique avec le tampon est fort, plus le signal attendu est intense et modulé.  $Q$  représente le vecteur de diffusion pour lequel

$$Q = 4\pi \sin\theta / \lambda, \text{ avec } \lambda \text{ la longueur d'onde}$$

Les informations apportées donnent accès à la forme et la taille de la macromolécule, et non pas à la structure secondaire ou tertiaire de cette dernière.

## ii) **Acquisition des données SAXS**

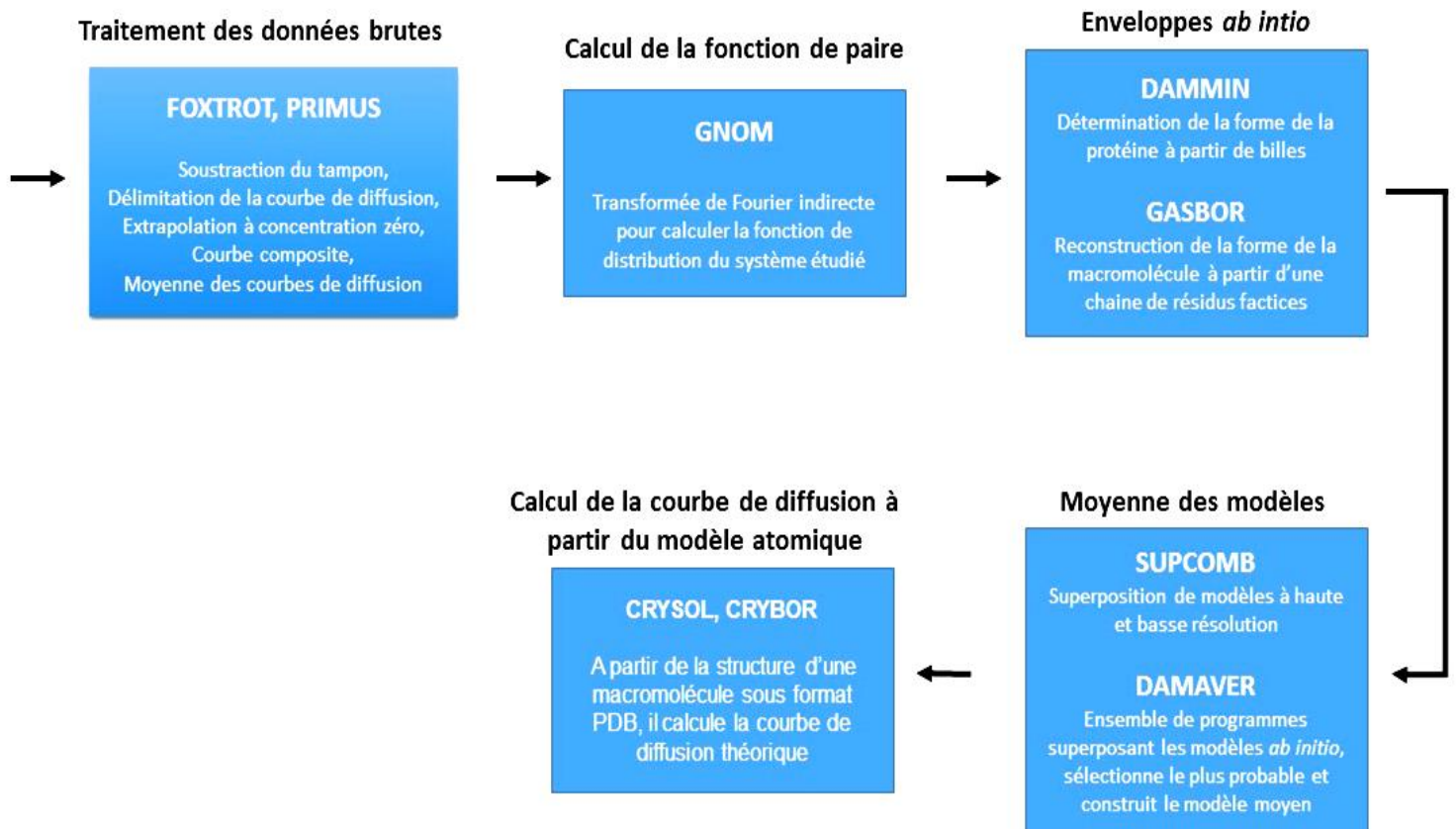
Les expériences SAXS ont été réalisées sur la ligne SWING au synchrotron SOLEIL (Saclay).

Les expériences ont été réalisées en passeur dans le tampon de la protéine même soit 25mM Tris pH8, 500 mM NaCl, 10% glycerol, 2mM DTT, 1mM EDTA. Ce tampon a été mesuré avant et après chaque injection de protéine. Trois injections de 40 $\mu$ L de la construction 26-1118 à ont été réalisées à 0,69mg.mL<sup>-1</sup>, 1,10 mg.mL<sup>-1</sup> et 2,35mg.mL<sup>-1</sup>. Les échantillons ont été injectés dans cet ordre afin d'éviter au maximum l'encrassement de la cellule de mesure. Les échantillons ont été injectés à 50 $\mu$ L/min dans la cellule de mesure thermostatée à 15°C. Dans ces conditions, nous avons pu exposer chaque échantillon entre 17 et 25 fois 1000ms.

## iii) **Traitement de données**

Le traitement des données a été effectué avec le logiciel Foxtrot développé à SOLEIL, notamment pour l'examen des courbes de diffusion et la soustraction de la courbe de diffusion du tampon. Les données ont été analysées par la suite ATSAS 2.6.0 développée à l'EMBL Hamburg. L'organigramme ci-dessous retrace et résume la stratégie adoptée afin d'aboutir à la forme globale des protéines en solution<sup>123</sup>.





- Traitement des données brutes

### Courbe de diffusion

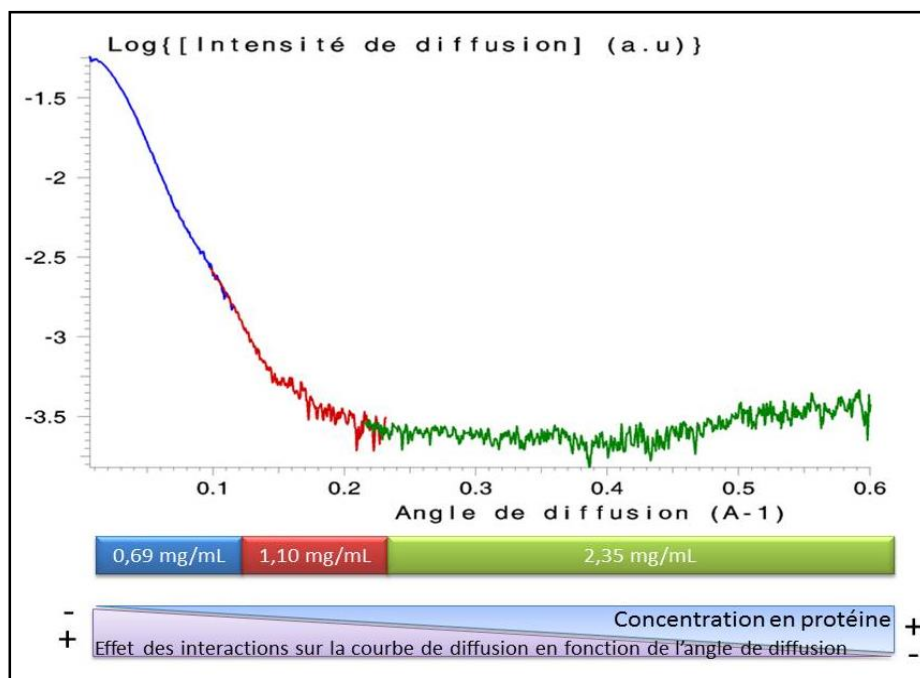
La courbe de diffusion du tampon a été retranchée à la courbe de la diffusion de l'échantillon pour obtenir la contribution de la protéine seule par le programme Foxtrot développé sur la ligne SWING du synchrotron Soleil :

$$I_{\text{protéine}}(Q) = I_{\text{échantillon}}(Q) - I_{\text{solvant}}(Q)$$

A partir de cette courbe de diffusion, il est désormais possible d'évaluer la validité des données en déterminant l'intensité à l'origine  $I_0$  et le rayon de giration  $R_g$  de la protéine à partir de la loi de Guinier. Pour cela il a été utilisé PRIMUS<sup>124</sup>, logiciel de la suite ATSAS. Dans notre cas, il y a un effet de concentration notable : le  $R_g$  calculé augmente avec la concentration. Il nous a donc fallu nous affranchir du facteur de structure (interactions intermoléculaires dépendant de la concentration).

Pour cela nous avons réalisé une courbe composite avec

- ⇒  $0,0086 < Q < 0,09 \text{ \AA}^{-1}$  pour la concentration la plus faible qui est moins polluée par les interactions
- ⇒  $0,09 < Q < 0,23 \text{ \AA}^{-1}$  pour la concentration intermédiaire
- ⇒  $0,23 < Q < 0,58 \text{ \AA}^{-1}$  pour la concentration la plus haute. Les grands angles étant moins impactés par le facteur de structure, nous pouvons nous permettre d'utiliser la courbe à plus haute concentration afin d'obtenir un peu plus de précision quant à la forme de la protéine.



**Figure 34-** Courbe composite utilisée pour le traitement des données

### Loi de Guinier<sup>125</sup>

Proche des angles  $Q=0$ , l'intensité diffusée d'une particule peut être décrite par une gaussienne. La validité du domaine dépend de la taille de la particule et est compris entre  $0 < Q \cdot R_g < 1,3$ . La loi de Guinier est équivalente à une variation linéaire de  $\ln[I(Q)]$  en fonction de  $Q^2$  :

$$\ln(I(Q)) = \ln(I(0)) - \frac{Q^2 R_g^2}{3}$$

Tous les points de la courbe de diffusion en amont du premier point utilisé par le Guinier peuvent être éliminés, dans une limite de  $q_{min} \cdot R_g < 1,3$ . Plus le Guinier débute tardivement, plus les données sont pauvres. A partir de la courbe tracée, nous avons accès au rayon de giration, qui dépend de la forme et de la taille de la protéine.

Il est défini par l'équation suivante :

$$Rg^2 = \frac{\int_V r^2 (\rho_{prot}(\vec{r}) - \rho_{tampon}) d\vec{r}}{\int_V (\rho_{prot}(\vec{r}) - \rho_{tampon}) d\vec{r}}$$

pour laquelle  $r$  est la distance entre le centre de masse de la protéine et chacun des atomes la composant. Il dépend donc de la taille et de la forme de la protéine.

### Poids moléculaire et loi de Porod

Indépendamment de la loi de Guinier, il est possible de déterminer le volume des particules hydratées  $V_p$  grâce à l'équation de Porod :

$$V_p = \frac{2\pi^2 I(0)}{Q}$$

Où  $Q$  : invariant de Porod, dépendant de  $I(s)$ . Typiquement, pour une protéine globulaire,  $V_p$  (en  $\text{nm}^3$ ) vaut 1,5 à 2 fois le poids moléculaire. Le logiciel en ligne Saxs Mow<sup>126</sup> utilise cette propriété afin d'évaluer celui de la protéine en solution et ainsi nous apporte une vérification quant à la qualité des données collectées.

### Degrés de compaction des particules : graphique de Kratky

Le graphique de Kratky  $Q^2 I(Q) = f(Q)$  donne des indications sur le degré de compaction de la protéine. Ainsi, une courbe en forme de cloche correspond à une particule globulaire alors qu'une protéine plus allongée donnera un plateau aux angles élevés. Cependant, il a été démontré par Bernardo et al (2009)<sup>127</sup> que la flexibilité n'est pas directement liée au taux de compaction de la protéine annoncé par ce graphique (une protéine annoncée compacte par le Kratky peut cacher une protéine flexible).

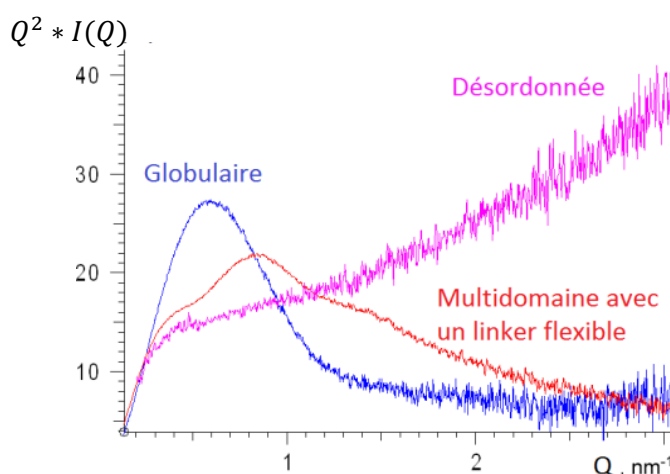


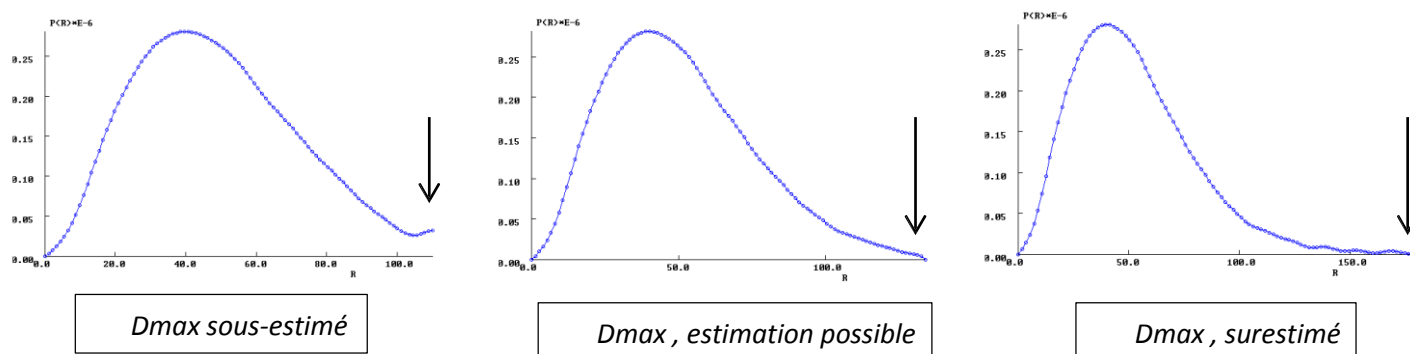
Figure 35 (adapté de 123)- Graphique de Kratky et interprétation possible

### Détermination de la fonction de distribution de paire

La transformée de Fourier directe de  $I(Q)$  mène à la fonction de distribution de paires définie par :

$$p(r) = \gamma_{\text{objet}}(r) \cdot r^2$$

où  $\gamma(r)$ , la fonction de corrélation, est un histogramme des distances entre chaque paire d'éléments qui diffusent dans la particule. Cette fonction de paires est donc proportionnelle au nombre de voisins à une distance  $r$  et est ainsi dépendante de la forme de la protéine. Par exemple dans le cas de particules multi-domaines, la courbe peut présenter des oscillations et des épaulements correspondant aux distances intra et inter sous-unité. Le calcul direct de la fonction de paires à partir de  $I(Q)$  est rendu difficile et risqué par les coupures en résolution  $Q_{\min}$ ,  $Q_{\max}$  et le bruit de la courbe de diffusion. Pour pallier à ce problème, le programme GNOM<sup>126</sup> permet de calculer la transformée de Fourier indirecte de  $I(Q)$ . Pour cela l'utilisateur doit définir le diamètre maximal  $D_{\max}$  de la protéine grâce aux graphiques  $p(r)=f(r)$  calculés par GNOM :



La fonction de distribution de paires peut aussi mener à  $I_0$  et  $R_g$ , cependant la sensibilité aux interactions ou à la faible présence d'oligomères est plus faible. Ces valeurs peuvent tout de même s'avérer utiles pour recouper les données calculées à partir du Guinier.

#### ▪ Construction de l'enveloppe de la protéine *ab initio*

A partir des fonctions de distribution de paires obtenues peut être modélisée l'enveloppe de la protéine. Pour cela deux logiciels aux stratégies bien différentes ont été utilisés : DAMMIN<sup>128</sup> et GASBOR<sup>129</sup>.

## DAMMIN

Sa stratégie générale est basée sur une méthode d'essai-erreur grâce à un Monte-Carlo couplé à un recuit simulé partant d'une sphère de diamètre Dmax imposé composée de billes.

Afin de retrouver la forme de la protéine, chacune des billes est aléatoirement considérée comme faisant partie du solvant ou bien de la particule. A chaque étape du recuit simulé, la configuration X d'une bille choisie aléatoirement est changée dans l'optique de minimiser la fonction suivante :

$$f(x) = \chi^2 \cdot [I_{\text{exp}}(Q), I(Q, X)] + \alpha P(X)$$

où  $\chi$  est la divergence entre la courbe de diffusion calculée et expérimentale, et  $P(X)$  est la pénalité assurant un résultat compact et non discontinu. Une contrainte de volume est imposée dès le départ afin que la protéine n'excède pas le Dmax choisi. Cette stratégie nécessite une délimitation de la courbe de diffusion aux grands angles de  $Q_{\text{max}} = 8/R_g$ . Il a été utilisé ici une limite de  $Q = 0,016 \text{ \AA}^{-1}$ .

Afin de réaliser un modèle fiable, 20 enveloppes ont été générées par DAMMIN. Afin de réduire l'ambiguïté de l'interprétation des données SAXS, un modèle moyen peut être conçu. Pour cela, SUPCOMB<sup>130</sup> superpose les différentes enveloppes et en définit le degré de similarité NSD. Ce NSD est basé sur la distance entre un point du premier modèle et tous ceux en son voisinage appartenant au second modèle et vice-versa. Ces distances sont additionnées et normalisées contre les distances entre les points voisins pour les deux enveloppes superposées. Après superposition de ces enveloppes, DAMAVER<sup>132</sup>, une suite de programmes, moyenne ces modèles et en garde le motif le plus persistant constituant ainsi l'enveloppe moyenne représentative de tout l'ensemble.

## GASBOR

Chacun des acides aminés de la protéine est représenté par un pseudo résidu et forme une seule chaîne solidaire. La densité électronique attribuée à chaque pseudo-résidu est la moyenne des densités électroniques des résidus dans la séquence de la protéine. Grâce à ce calcul, il est attribué à chaque résidu un pouvoir diffusant moyen dans l'eau.

Ici encore une minimisation par recuit simulé est utilisée. Une pénalité assure un résultat compatible avec la forme que peut adopter une chaîne polypeptidique.

Pour cela, l'histogramme de distance moyenne entre voisin calculé doit être similaire à celui d'une protéine globulaire. Il est aussi retrouvé la contrainte de volume imposée par le Dmax.

## 5) Cristallogénèse

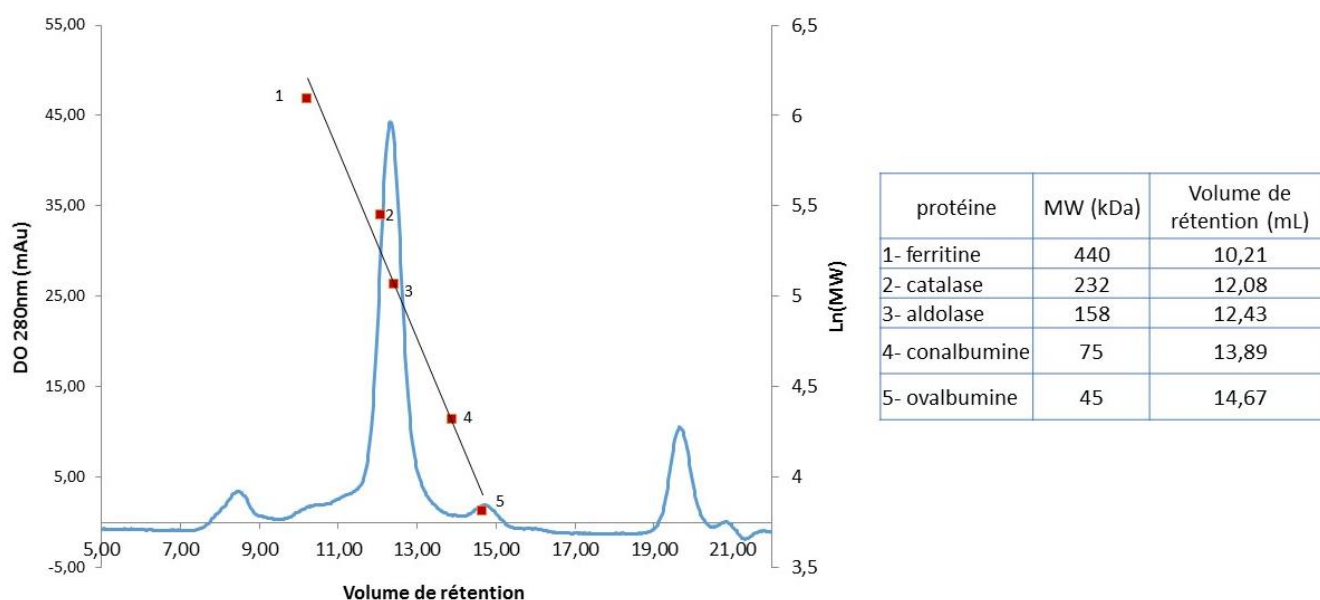
Un premier test de criblage de cristallogénèse a été réalisé par Inès Gallay, Ingénieur de recherche à l'IBBMC. Nous avons choisi des conditions classiques pour des protéines solubles (boîte The classics, QIAGEN). La boîte a été réalisée sur Mosquito à 20°C avec un échantillon de protéine concentré à  $3 \text{ mg.mL}^{-1}$ . Aucune piste n'a été obtenue.



## B. Résultats

Une des caractérisations de protéine effectuée en laboratoire consiste tout simplement aux études sur gel filtration. Dans notre cas, deux types de chromatographie d'exclusion ont été utilisés :

<i>Tamis moléculaire</i>	<i>Séparation</i>	<i>Type de résine</i>	<i>[NaCl]</i>	<i>[glycérol]</i>	<i>MW apparent</i>
<b>Superose 6 10/300 GL</b>	5-5000 kDa	Agarose hautement réticulé	500mM	10%	133 kDa
			100-200 mM	10%	Interaction forte
<b>S200 10/300 GL increase</b>	10-600 kDa	Agarose et dextran hautement réticulé	500mM-1M	10%	165,67 kDa

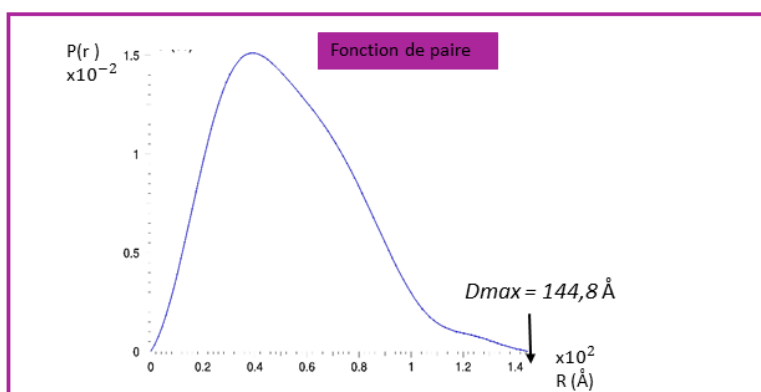
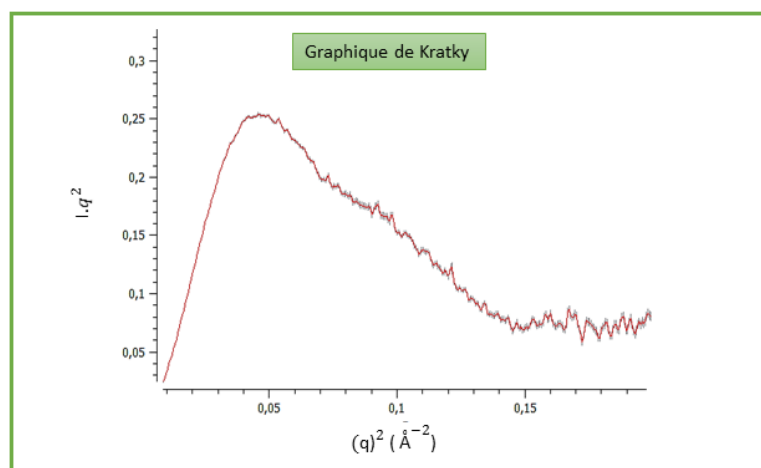
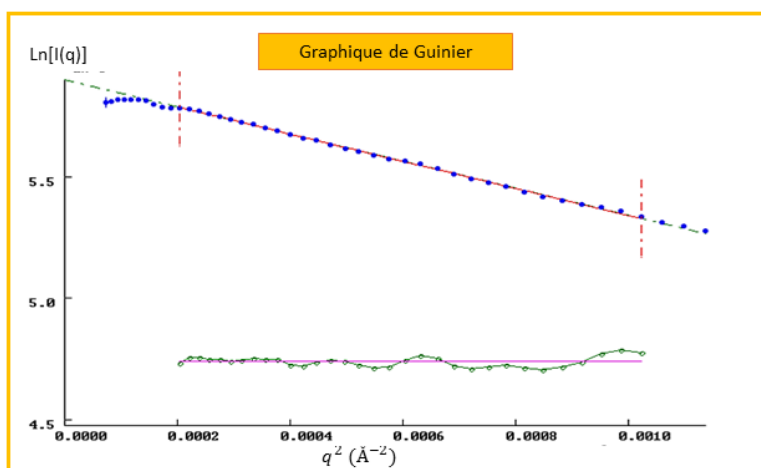


**Figure 36- Etude par gel filtration du module 26-1118.**

Dans le cas de l'étude sur Superose 6 10/300GL, le fragment 26-1118 interagit avec la matrice. En effet, en dessous de 300mM NaCl, la protéine d'intérêt ne ressort pas de la colonne. Or, cette interaction est aussi observée avec la résine de glutathion : difficulté à l'élution, concentration en glutathion augmentée à 50mM pour une élution optimale. Compte tenu de ces observations, notre fragment interagit avec l'agarose. Par ailleurs, à partir de 300mM NaCl, l'interaction avec la matrice de la Superose 6 10/300 GL est atténuée. Il existe donc des interactions polaires entre 26-1118 et la matrice de la colonne de chromatographie d'exclusion. La protéine est éluée à un volume qui correspondrait à une protéine globulaire de 133 kDa, soit environ son poids moléculaire théorique de 123kDa.

La colonne de chromatographie d'exclusion S200 10/300 GL increase possède quant à elle un revêtement formé en partie de dextran réticulé qui, théoriquement, empêche ces interactions hydrophobes. La protéine en ressort en un pic unique correspondant au poids moléculaire d'une protéine globulaire de 165 kDa. Or, 26-1118 a un poids moléculaire théorique de 123 kDa. Au vu de nos analyses bio-informatiques et de cette analyse par gel filtration, 26-1118 ne serait pas complètement globulaire. Ce résultat confirme une interaction entre la protéine et la matrice de la Superose 6 10/300GL. Nous pouvons conclure que 26-1118 est vraisemblablement monomérique et oblongue. Seules les expériences de SAXS ont permis une analyse précise de la forme du fragment protéique.

Dans un premier temps il a été nécessaire d'évaluer la qualité des données SAXS collectées et traitées. En effet, le SAXS est une technique puissante mais très sensible à la présence d'agrégats et aux interactions, qu'elles soient attractives ou répulsives. Il est ainsi très facile d'obtenir des résultats artefactuels. Globalement, les analyses démontrent des données SAXS de qualité qui semble raisonnable. Cependant, des analyses supplémentaires sur ces données, notamment le calcul de la masse moléculaire à partir de l'intensité à l'origine, sont à effectuées afin de savoir si nous nous sommes totalement affranchis de l'effet des interactions attractives sur la courbe de diffusion.



	MW (kDa)	$R_g$ (Å)	$I_0$ (unité exp)
Loi de Guinier (globulaire)	/	$39,4 \pm 8,03 \cdot 10^{-2}$	$365,99 \pm 0,5$
Loi de Porod	145,1	$41,53 \pm 0,065$	/
$P(r)$	/	41,53	366

Le *graphique de Guinier* est compris entre les angles de diffusions  $0,016 < q < 0,046 \text{ \AA}^{-1}$  avec  $0,586 < qR_g < 1,3$ , attestant de la fiabilité de la zone du Guinier. Les résiduels, colorés en vert en dessous du graphique de Guinier, sont faibles et semblent répartis de manière aléatoire autour de zéro. Primus estime une qualité de données à 91% . Ce calcul prend en compte les limites du Guinier en calculant  $Q_{\min/\max} \cdot R_g$ , le nombre de points écartés de la courbe de diffusion en amont du Guinier ainsi que la pertinence de la valeur de rayon de giration  $R_g$  calculée.

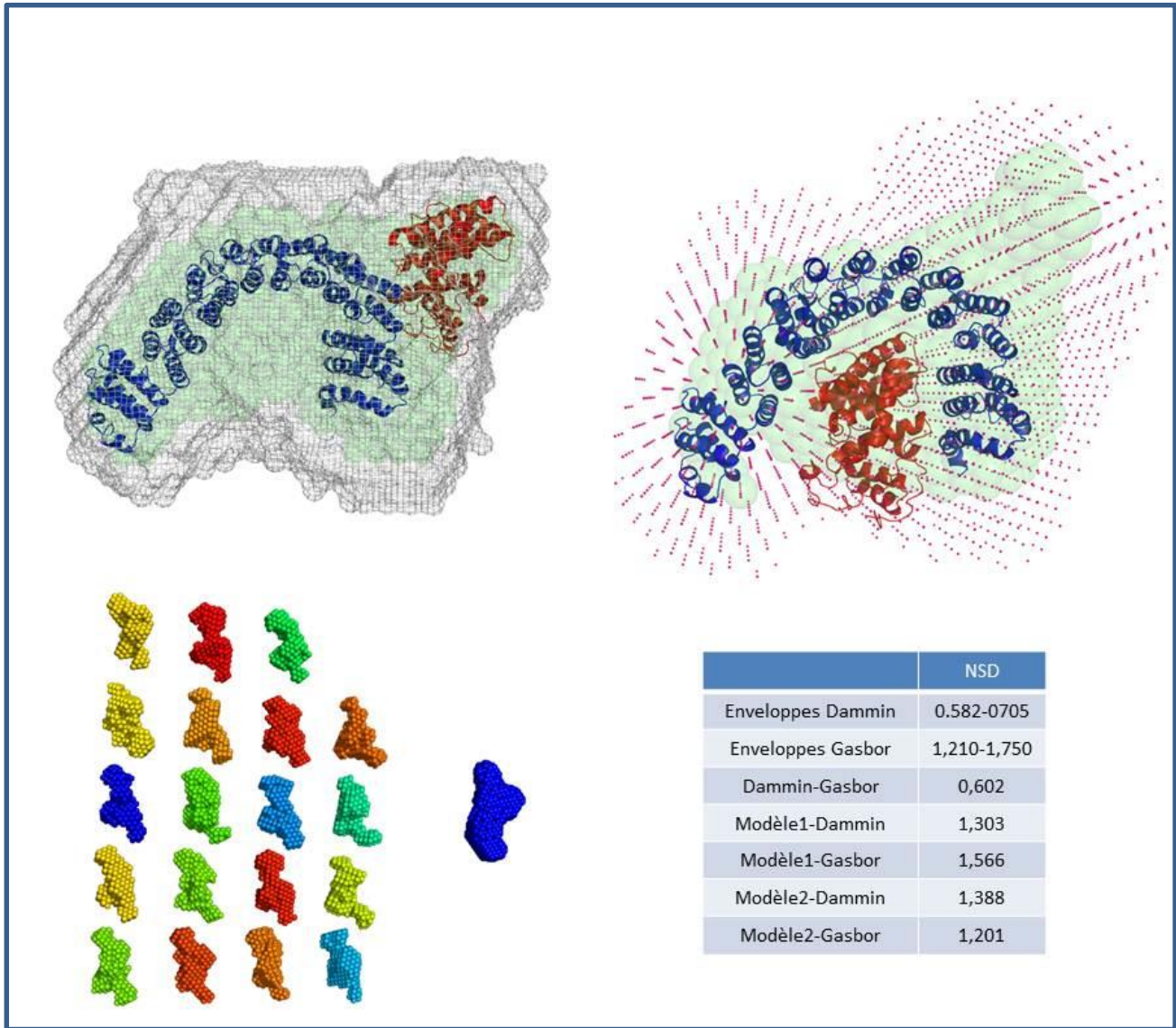
La *fonction de paire* est en forme de cloche non symétrique avec une faible modulation. Cette modulation témoigne de l'asymétrie de la protéine et confirme qu'elle n'est pas globulaire. De plus, la  $P(r)$  est un peu étirée vers la droite, signe que la protéine est allongée. Le diamètre maximal  $D_{\max}$  de  $144 \text{ \AA}$  n'est pas improbable pour une protéine de  $123 \text{ kDa}$  allongée. De plus, le rayon de giration  $R_g$  et l'intensité à l'origine  $I_0$  calculés restent proche de ceux estimés par la loi de Porod et de Guinier, signe de la bonne qualité des données.

La courbe obtenue par le *graphique de Kratky* est en forme de cloche avec un épaulement sur la droite. Par conséquent, le module N-terminal est probablement composé de deux domaines, comme nous l'avions proposé précédemment. De plus, la courbe a tendance à remonter pour les valeurs de  $q^2$  les plus élevés. Ainsi, la protéine comporte une partie flexible. Or, une région désordonnée avait été prédite entre les deux domaines potentiels.

Le poids moléculaire calculé à partir de la *loi de Porod* correspond à  $145.1 \text{ kDa}$ , soit  $20 \text{ kDa}$  de plus que le poids moléculaire théorique du module 26-1118. Cependant, ces valeurs sont tout de même très proches.

Enfin, le rayon de giration et l'intensité à l'origine calculés par ces différentes méthodes sont en accord, avec un écart n'excédant pas 5%.

Au vu de ces différentes analyses, les données traitées apportent des données compatibles avec nos analyses bio-informatiques et sont exploitables afin de générer une structure basse résolution.



**Figure 37-** Evaluation des enveloppes générées à partir des données SAXS par DAMMIN et GASBOR.

**En haut à gauche :** en gris, l'enveloppe moyenne de DAMMIN par DAMAVER ; en vert, l'enveloppe moyenne filtrée par DAMFILT,

**En haut à droite :** en rose, l'enveloppe moyenne de GASBOR par DAMAVER ; en vert, l'enveloppe moyenne filtrée par DAMFILT.

Pour ces deux enveloppes, les deux modèles I-TASSER sont positionnée l'un par rapport à l'autre suivant les deux possibilités : le domaine N-terminal en interaction avec la surface concave ou convexe du domaine à répétitions HEAT. en bleu le modèle I-Tasser de 553-1118 ; en rouge le modèle I-Tasser de 1-300.

**En bas à gauche :** les 20 modèles générés par DAMMIN

**En bas à droite :** Comparaison des enveloppes et modèles par superposition

Ces enveloppes ont été générées par DAMMIN et GASBOR, deux logiciels aux stratégies un peu différentes mais basées toutes deux sur le recuit simulé. Dans l'ensemble, les 40 enveloppes générées sont semblables (NSD proche de 1). Une différence un peu plus élevée est notable entre les enveloppes de GASBOR. Ceci s'explique par des enveloppes plus résolutivees générées avec des données utilisées jusqu'à la limite de résolution.

Ces enveloppes sont toutes allongées avec une protubérance globulaire. Les enveloppes moyennes de DAMFILT (enveloppe interne verte) et de DAMAVER (enveloppe externe) permettent de préciser la forme globale de la protéine qui est formée d'une partie centrale en forme de fer à cheval, et d'une excroissance compacte. Par ailleurs, ces deux enveloppes moyennes sont très similaires. Ainsi, la courbe de diffusion mène à une solution globale unique suivant un schéma précis. Cependant, le SAXS mène à des enveloppes moyennant les conformations adoptées en solution par la protéine. Nous pouvons donc affirmer que deux domaines composent 26-1118, mais pas que les domaines ne sont pas mobiles l'un par rapport à l'autre. Le  $\chi^2$  de chacune des enveloppes est raisonnable mais un peu élevé (entre 2,549 et 2,583). Au vu des résultats qui semblent fiables, nous avons essayé de superposer les modèles tridimensionnels du module N-terminal à l'enveloppe moyenne de DAMFILT.

Le modèle intégral du module 300-1118 n'est pas compatible avec les données SAXS. Cependant, la délimitation du domaine putatif composé de motifs HEAT avait été incertaine et la courbure du fer à cheval, le taux de compaction ainsi que la longueur des hélices ne sont pas des éléments facilement prédits. C'est pourquoi j'ai par la suite superposé le [modèle 553-1118](#) (en bleu sur la figure 37), plus fiable car issu d'un homologue. Le **domaine putatif compact N-terminal** (modèle rouge) peut être placé en deux endroits : au sein du domaine en fer à cheval, ou bien en interaction avec sa surface convexe. Cependant, la qualité de nos données SAXS et de nos modélisations ne nous permettent pas de définir l'emplacement du domaine putatif 1-300 par rapport au domaine en solénoïde  $\alpha$  ni même de vérifier le nombre de domaines composant le module 26-1118 (2 voire 3).

Ces expériences SAXS ont été réalisées dans un tampon riche en sel et avec une viscosité certaine. En effet, il contient 300mM NaCl et 10% de glycérol. Or, nous soupçonnions le glycérol et le sel de stabiliser la structure globale du fragment 26-1118. Afin de vérifier cette hypothèse, nous avons étudié la stabilité de la protéine par des expériences de thermofluor.

Le fragment 26-1118 est une protéine assez fragile, dont la dénaturation débute relativement tôt, autour de 35°C. Le pic unique de déaturation est cependant en partie en désaccord avec nos prédictions de solénoïde  $\alpha$ . En effet, il a été prouvé que plus de tels domaines possèdent de répétitions, plus le pic de dénaturation est asymétrique et fin<sup>133</sup>. Or, la courbe de dénaturation obtenue est asymétrique mais large.

Il a été proposé un modèle de dénaturation pour les domaines à motif hélicaux répétés, basé entre autre sur des études calorimétriques et le modèle d'Ising. Entre la population initiale repliée et la population finale dénaturée, il existerait un ou des états de transition aux alentours du Tm pour le(s)quel(s) la protéine serait partiellement dénaturée en N- et C-terminal. Cependant, la comparaison des thermogrammes obtenus pour les protéines globulaires et motifs répétés ne permet pas de mettre en évidence ces états de transition<sup>132,133</sup>.



La largeur du pic ainsi que la faible température de demi-dénaturation observées dans le cas de 26-1118 peuvent s'expliquer par l'absence du module C-terminal de PI4KA ou d'un partenaire démasquant entre autre certaines régions hydrophobes du module.

Le criblage des différents tampons peut se diviser en quatre grands ensembles observables sur la figure 38 :

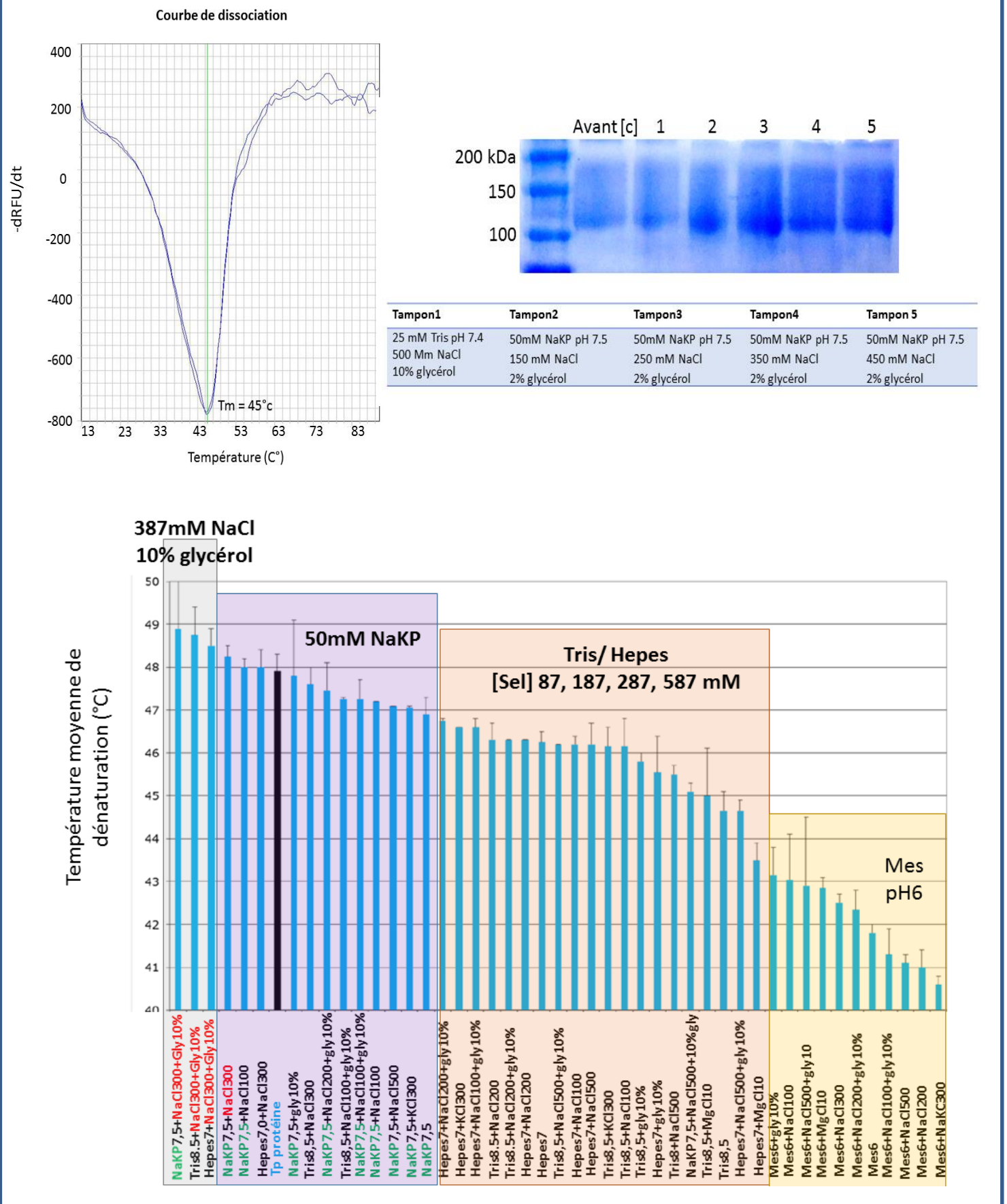
Les conditions les plus favorables à la stabilité de la protéine sont composées de 387mM de NaCl et 10% de glycérol qui est donc bien une molécule ayant un effet sur la stabilité du fragment 26-1118. La nature du tampon utilisé ne semble pas importante. Cependant le pH reste dans une gamme physiologique comprise entre pH 7 et 8.

Le phosphate sous forme de NaKp 50mM ph 7,4 semble stabiliser énormément la protéine, que ce soit en présence de glycérol ou non, avec une concentration en NaCl comprise entre 100 et 500mM. Il est connu que le phosphate stabilise les protéines en général.

S'en suit les solutions tamponnées en Hepes et en Tris, sans grandes distinctions entre les conditions. Cependant, cet ensemble englobe la solution NaKP ph7,5 , 587 Mm NaCl, 10% glycérol, qui abaisse le Tm du fragment à 45°C. Il semblerait donc que de fortes concentrations en sel affectent la stabilité de la protéine au-delà des effets bénéfiques du phosphate.

Enfin, des pH plutôt acides aux alentours de 6 en MES sont tout à fait délétères sur la protéine.

Les conditions les plus favorables sont tout de même très proches au niveau des TM mesurés. Nous avons donc tenu à vérifier à quel point la concentration en sel permettait la stabilisation du module N-terminal. Pour ce faire, des tests de concentration ont été réalisés avec des tampons de compositions différentes répertoriées dans le tableau de la figure 38. Nous avons essayé de diminuer la concentration en glycérol afin d'optimiser l'étape de fixation à la résine. Le tampon optimal contient 250 mM NaCl, ce qui correspond au tampon optimum sans glycérol. De plus, un fort effet du phosphate sur la solubilité du fragment est observé, comme attendu.



**Figure 38-** Etude de la stabilité du fragment 26-1118 de la PI4KA humaine par thermofluorescence et tests de concentration

**En haut à gauche** : Courbe de dénaturation de la protéine représentant l'opposé de la dérivée permettant d'estimer la température de demi-dénaturation. Dans le tampon initial, 26-1118 a une température de demi-dénaturation  $T_m$  de 45°C.

**En bas** : Test screen de tampon Tris, Phosphate, Hepes et Mes. La figure présente les valeurs moyennes des températures de dénaturation. Les valeurs sont classées en ordre décroissant.

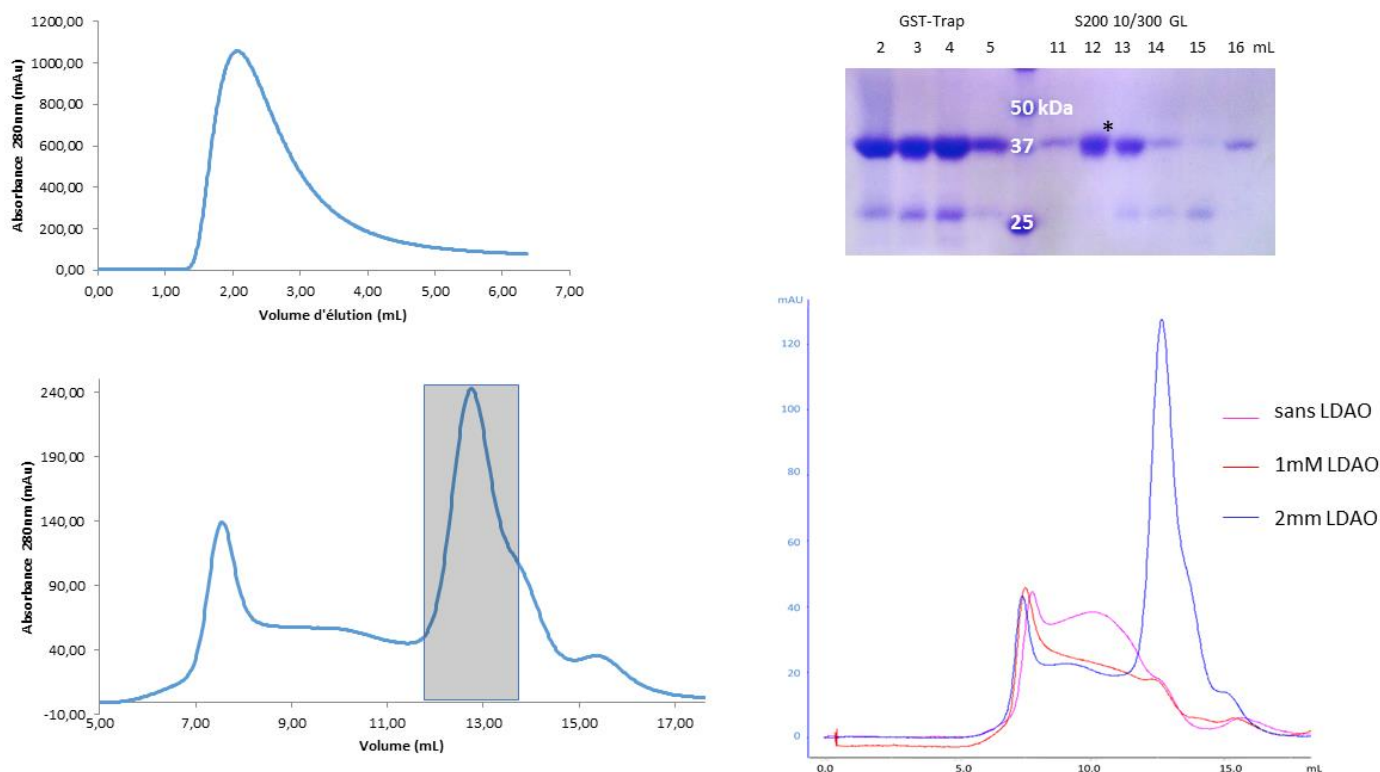
**En haut à droite** : Tests de concentration de 26-1118 et différents tampons testés. Le sel entre 150 et 250 mM NaCl est bénéfique à la bonne concentration de la protéine en présence d'un faible pourcentage de glycérol. Ce dernier peut donc être retiré des solutions utilisées pour la chromatographie d'affinité.

## II. Caractérisation du domaine putatif DI

---

### **A. Article n°3: A new potential for N,N-Dimethyldodecylamine-N-oxide surfactant based on the Amber force field for Molecular Dynamics Simulations of Membrane Proteins and application to a fragment of the lipid kinase PI4KA**

La seconde construction que nous avons décidé d'exprimer et de purifier à grande échelle est le domaine putatif DI. En effet, au vu des résultats obtenus par nos collaborateurs, il s'agit du domaine le plus susceptible d'interagir avec le domaine d1 de NS5A. Lors de la mise au point de la purification, nous nous sommes rapidement rendu compte que la protéine avait besoin de détergent au-dessus de la CMC pour être soluble (voire figure 39). Nous avons choisi d'utiliser le N, N-dimethyl-dodecylamine oxide (LDAO), un détergent zwitterionique utilisé pour la solubilisation de protéines membranaires intégrales et leur cristallisation<sup>134</sup>. Au contraire du Triton X-100, le LDAO n'absorbe pas à 280nm, permettant ainsi une meilleure estimation de la quantité de protéine purifiée ainsi qu'un meilleur suivi de la purification. La purification de deux protéines de fusion a été mis au point: 6his-DI, possédant une étiquette hexa-histidines en N-terminal; et GST-DI, avec une étiquette GST. Elles sont toutes deux purifiées suivant le même protocole : chromatographie d'affinité suivie d'une chromatographie d'exclusion (**figure 39**). A noter que DI, entre nos mains, absorbe toujours à 260 nm. Cependant, aucun autre test portant sur la détection d'acides nucléiques n'a été effectué lors de la caractérisation de ce fragment. La mise au point de la purification de DI a été l'occasion de former Anaëlle Blöch, deuxième année de licence Bioplus à l'Université Paris-Sud XI, aux techniques de clonage, d'expression et de purification des protéines.



**Figure 39-** Suivi des différentes étapes de purification de GST-DI.

**En haut à gauche :** chromatogramme enregistré à 280nm lors d'élution de GST-DI.

**En bas à gauche :** chromatogramme enregistré à 280 nm de la chromatographie d'exclusion Superdex 200 10/300GL.

Le pic correspondant à la construction GST-DI est grisé. Un épaulement sur la droite du pic correspond à de la GST seule.

**En haut à droite :** gel Tris-Tricine 14% correspondant aux deux étapes de purification de la protéine d'intérêt.

**En bas à droite :** figure tirée de l'article n°3. Trois expériences de chromatographie d'exclusion ont été réalisées : sans LDAO, avec LDAO sous et au-dessus de la CMC. Ainsi, nous pouvons constater que la protéine a besoin de détergent au-dessus de la CMC pour ne pas agréger.

En parallèle, Stéphane Abel, ingénieur-chercheur au CEA Saclay, et Esra Karakas, doctorante, ont mis au point une nouvelle paramétrisation des charges de la molécule de LDAO. Afin de valider cette paramétrisation, des simulations de dynamique moléculaire avec le LDAO seul sous le champ de force AMBER ont été réalisées. Ceci a permis de vérifier que le système permet la formation de micelles pures conformes aux caractérisations expérimentales publiées. Il s'est donc agi dans un second temps de valider ce système avec du LDAO en présence de protéine. Pour se faire, nous avons monté une collaboration sur GST-DI. Afin de mener des simulations de dynamique moléculaires fiables, il a fallu dans un premier temps modéliser DI. N'ayant pas d'homologues de structure connue de cette protéine, des modèles ont été générés par threading grâce à I-TASSER et QUARK. Le meilleur modèle obtenu pour chacune de ces modélisations a été utilisé pour les simulations de dynamique moléculaire avec LDAO. Ces deux modèles, sur des simulations de dynamique moléculaire de 200 ns, sont globalement stables et donc potentiellement possibles.

Seule l'hélice N-terminale, pour laquelle les prédictions de structure secondaires ne trouvent pas de limite bien définies, est un peu moins stable que le reste des modèles.

Simultanément, DI a été caractérisée tout d'abord par SEC-MALLS. Nous avons pu ainsi évaluer à 41 le nombre de molécules de LDAO interagissant avec un monomère de DI sans étiquette. Cependant, la variation de l'indice de réfraction était trop faible pour calculer le rayon hydrodynamique du complexe.

Dans un second temps, nous avons étudié l'enfouissement des 4 tryptophanes de DI par fluorescence des tryptophanes. Deux des tryptophanes (1234 et 1236) sont séparés d'un demi-pas d'hélice  $\alpha$  et on attend que l'un soit exposé au solvant. Nos études de fluorescence ont permis de mettre en évidence son interaction avec le LDAO puisque les quatre tryptophanes sont dans un environnement hydrophobe, avec un maximum de spectre d'émission à 331 nm. A noter que le Trp 1234 est conservé entre PI4KA et que le Trp 1236 est conservé entre PI4KA et STT4p. Les deux autres tryptophanes, quant à eux, ne le sont pas.

Les résultats obtenus par simulations sont en accord avec nos données, cependant des études sur une protéine de structure connue seraient un avantage certain quant à la validation de ce système.





# A new potential for N,N-Dimethyldodecylamine-N-oxide surfactant based on the Amber force field for Molecular Dynamics Simulations of Micelles and application to a fragment of the lipid kinase

## PI4KA

*Esra Karakas<sup>1,2,3‡</sup>, Cyntia Taveneau<sup>3‡</sup>, Stéphane Bressanelli<sup>3</sup>, Massimo Marchi<sup>2,3</sup>, Bruno Robert<sup>2,3</sup> and Stéphane Abel<sup>2,3\*</sup>*

<sup>1</sup>Maison de la Simulation, USR 3441, CEA – CNRS – INRIA – Univ. Paris-Sud – Univ. de Versailles, 91191,

<sup>2</sup>Commissariat à l'énergie atomique et aux énergies alternatives, Gif-sur-Yvette Cedex, France DSV/iBiTec-S/SB2SM/LBMS & CNRS UMR 9198

<sup>3</sup>Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Université Paris-Sud, 1 avenue de la terrasse, 91198 Gif-sur-Yvette, France

‡ Equal contributions

\* Corresponding authors: [stephane.abel@cea.fr](mailto:stephane.abel@cea.fr) & [stephane.bressanelli@vms.cnrs-gif.fr](mailto:stephane.bressanelli@vms.cnrs-gif.fr)

## Abstract

In this paper, we describe the development and the validation of original atomic partial charges (RESP atomic charges) for the N, N-dimethyl-dodecylamine oxide (LDAO) surfactant. These charges were designed to be fully compatible with all the AMBER force fields for proteins. These new charges were validated with molecular dynamics (MD) simulations of two different systems: pure LDAO micelles and a LDAO-protein complex. In the first case, we used different AMBER force fields (i.e. Amber99SB-ILDN, Amber12SB and GAFFLipid) and starting conditions to examine different characteristic micellar properties (such as shape, size, hydration and LDAO headgroup properties) against experimental and theoretical results. We show that the micelle properties do not significantly depend on the simulation protocol and the force field parameters and are in good agreement with experiments. In a second test case, we examine the ability of our parameters to model a LDAO-protein complex formed with a fragment of the lipid kinase PIK4A modeled with the I-TASSER and QUARK molecular modelling servers. This fragment (DI) was purified in the presence of LDAO and characterized by SEC-MALLS (size exclusion chromatography-multi angle laser light scattering) and fluorescence spectroscopy. These experiments show that 41 LDAO molecules associate with each DI and that all 4 DI tryptophans are in apolar environments. We carried out ~200 ns MD simulations with each DI model and 41 LDAO molecules modelled with the Amber99SB-ILDN force field. The simulations show different results as function of the model, in particular in the LDAO binding region, the secondary structure of the two models and the accessibility to the solvent of the 4 tryptophans and we find that only the QUARK model agrees well with the experimental data.

## I. Introduction

In addition to being routinely used by pharmaceutical industry and cosmetics, surfactant molecules are also important in the investigation of membrane proteins. Indeed, since surfactants mimic a membrane environment, they are used to solubilize and to keep membrane proteins biologically active. Among these surfactants, N,N-dimethyl-dodecylamine oxide (LDAO, Figure 1) is an amphiphilic surfactant with a high hydrophilicity character [1][2] and has been widely used in the past [3]. In an aqueous environment, depending on the experimental conditions, LDAO can form different meso-structures such as micelles or lamellar phases [4]. Above the critical micelle concentration (CMC) LDAO forms micelles. To the best of our knowledge, the first work that examined the structural properties of LDAO micelles was done by Herrmann et al. in the early 80's [5] with static light scattering experiments. In this work, the authors showed that 76 LDAO monomers are needed to form a stable micelle. More recently, other groups reported aggregation numbers in the range of 67 - 110 molecules [6][7][8][9][10]. In more recent small angle neutron scattering (SANS) experiments, Barlow et al.[11] found an aggregation number around 104. The large variations of the aggregation numbers estimated in experiments can be explained by the fact that inter-particle interactions were not taken into account or the presence of ions that lead of a partial ionization of the LDAO headgroup [6].

Molecular dynamics (MD) simulations have been used in the past to gain insights in the structures of surfactant assemblies, in particular, Marchetti [12] and Sterpone et al. [13] examined the structure and dynamics of water in the solvation shell of a model of LDAO micelles. LDAO was modeled with the CHARMM22 [14] force field and its charges were computed from density functional theory (DFT) calculations. More recently, Lorenz et al. [15] developed an OPLS-AA force field [16] to examine the structural properties of micelles containing LDAO alone and LDAO in association with a mixture of C4/C8 ethyl ester oils. All these authors, using an aggregation number of 104 [11], found that the LDAO micelles had an

ellipsoidal shape with a size of around 19 Å and average hydration number per LDAO around 8, in good agreement with experiments [7][11]. Although the AMBER force field is widely used to simulate membrane proteins (see for instance refs [17][18]), to the best of our knowledge, no AMBER parameter sets are available for LDAO. To fill the gap, in this paper we have derived atomic partial charges (or Restrained ElectroStatic Potential, RESP) for LDAO with a quantum theory compatible with AMBER force field for proteins and its successive derivations. As previously applied to other surfactants [19] [20], we used a “building block” approach to control the charge derivation step in a way that is consistent with the Cornell et al. method [21] for peptide fragments. To test the transferability of the computed RESP charges and validate our approach, we used two test cases. In the first test case, we simulated a model of pure LDAO micelle, where the RESP charges were combined with three AMBER force fields (i.e. Amber99SB-ILDN [22][23], Amber12SB [24] and GAFFLipid [25] (based on the General Amber Force Field [26])). The only difference between these three force fields in case of LDAO is for the CCCC torsion and Lennard-Jones parameters which were changed to better reproduce the long alkyl chain properties.

In the second case, we simulated a model of protein-LDAO complex formed with a fragment of the human lipid kinase PI4KA called through the paper “DI” and comparing to experiments described in this work and in a forthcoming paper (Taveneau, Blondeau & Bressanelli, submitted). PI4KA is a 2102-residue protein that transiently associates with the cytosolic leaflet of membranes to perform its phosphoryl transfer activity to the headgroup of phosphatidylinositol [27]. We previously reported that the first 1100 N-terminal residues of PI4KA are dispensable for this function and particularly for the support of hepatitis C virus (HCV) replication [28]. By sequence analysis, we also defined a likely pivotal domain of some 130 residues, which we term PI4KA “DI”, located right after the 1100 N-terminal residues. We further verified that DI is part of the functional C-terminus of PI4KA and thus potentially involved in membrane recruitment (Taveneau, Blondeau & Bressanelli, submitted). In this



work, we show that the use of LDAO is essential to solubilize and characterize the DI protein structure. As the DI protein three-dimensional structure is not known, we construct for this work two homology models for DI with the I-TASSER [29] and QUARK [30] web servers and use MD simulations to compute a few biophysical properties of the protein to compare with experimental results from LDAO-solubilized purified DI.

The paper is organized as follows: In section II, we describe the derivation of the RESP charges for the LDAO, the construction of our micellar systems with two approaches called “preformed” (PF) and “self-assembled” (SA), the obtaining of our two *ab-initio* models with I-TASSER and QUARK servers and finally, the experimental details for the LDAO-solubilized DI purification and characterization by tryptophan fluorescence and size exclusion chromatography-multi angle laser light scattering, SEC-MALLS). In section III, we present our results on different micellar characteristic properties (such as micelle size, shape, surfactant conformation and hydration) and, finally, in section IV, we describe the interactions between the PI4KA DI models and the LDAO micelle.

## II. Material and Methods

### II.1 RESP Charge Derivation

The RESP charges derivation for LDAO was carried out by using the “building blocks” approach [19][20] where the LDAO molecule was split into two elementary chemical blocks with well-defined and controlled conformations (here, *all-trans*) [31]. As shown in Fig. 2 we chose to split LDAO into the N,N-dimethyl amine oxide headgroup and the N-alkyl chain tail. The geometry optimization, frequency calculations and molecular electrostatic potential (MEP) computations of each building block were carried out in the gas phase with the Gaussian 2009 program [32] using the HF/6-31G\* level of theory [33]. This level of theory in the MEP computation leads to an implicit polarization required for condensed phase simulations with the AMBER force fields [34]. Only the building block conformations with the lowest minima

were involved in the charge derivation procedure. The charge fitting was performed in two steps with hyperbolic constraint values of 0.0005 and 0.001 with the RESP program [34] (version 2.2 available at <http://q4md-forcefieldtools.org/RED/resp/>). The Connolly algorithm was used to compute the grid point involved in the MEP calculation [35]. To control the molecular orientations of the optimized molecules, we applied the rigid-body orientation algorithm implemented in the RED IV program leading to reproducible charge values [31]. Two molecular orientations based on the C1, N and C2, and C3, C4 and C5 sets of atom names for N, N, N-trimethylamine-N-oxide and the alkyl chain (Figure 2) were involved in charge derivation. To keep the total group charge to zero, different inter-molecular charge constraints, involving the terminal methyl groups of the N, N, N-trimethylamine-N-oxide and the C5-C15 n-alkyl chains, were applied during the charge fitting step to construct the LDAO surfactant and analogs. The RRMS (Relative Root Mean Square) value between MEP values calculated by quantum mechanics, and those generated using the derived charge values was 0.3142. To automatize the RESP charge derivation procedure, we used the RED server (<http://q4md-forcefieldtools.org/REDS>)[36].

To evaluate the transferability of the LDAO RESP charges to different AMBER force fields, we combined them with 3 AMBER force fields (i.e. Amber99SB-ILDN [23], Amber12SB [24] and GAFFLipid [25]). We point out that the bonded and non-bonded parameters for the N-O group not being included in the Amber99SB-ILDN and Amber12SB force fields, the GAFF parameters were added to those potentials for the corresponding atoms (i.e. N4 and O atom types for the N and O atoms, respectively). The atom types as well as the computed RESP charges used to model the LDAO molecule are listed in Table S1 of the supporting information.

## II.2 Simulation setup

### II.2.1 Test case 1: Simulation of a pure LDAO micelle

To compare our simulation results with previous simulation results, we chose to construct a micelle of 104 LDAO monomers. The surfactant were solvated in a cubic box of 100 Å of side containing 32 000 TIP3P [37] water molecules. To examine the influence of the simulation initial conditions, the “preformed” (PF) and the “self-assembled” (SA) protocols were used to generate the initial system coordinates. In the PF approach, a spherical micelle is constructed with the Packmol program [38]. This program builds an initial MD starting point by packing the surfactant molecules in defined regions of space forming, thus forming spherical micelle. We used two spheres in this construction: Initially, the hydrophobic end of LDAO was oriented inside a first sphere with the C1 carbon of the alkyl chain (see Fig. S1 in the Supporting Information) placed on the outer sphere of 17 Å, whereas C12 was placed inside an inner sphere of radius of 4.0 Å. SA simulations, instead, were constructed by randomly placing the 104 LDAO monomers within the simulation box.

The simulations presented in this paper were carried out with GROMACS (v4.5.5 and 4.6.5) MD package [39]. To prepare the micellar systems prior to the simulation production runs, each micellar system was minimized with the steepest descent algorithm and an energy tolerance (energy difference between two consecutive steps) lower than 1000 kJ.mol<sup>-1</sup>.nm<sup>-1</sup>. In the second stage, the minimized systems were equilibrated in the NVT ensemble at 298 K for 400 ps by using the Berendsen thermostat [40] with a coupling constant  $\tau_T = 0.1$  ps. The resulting molecular systems were then equilibrated in the NPT ensemble at ambient conditions (T = 298 K and P = 1.015 bar) for 500 ps using the Bussi et al. thermostat [41] ( $\tau_T = 0.1$  ps) and the Parrinello-Rahman barostat [42][43] ( $\tau_P = 3.0$  ps), with a compressibility value of 4.5.10<sup>-5</sup> bar<sup>-1</sup>. The LDAO and the water were separated in two thermostat baths. During the NVT and NPT equilibration steps, LDAO molecules were harmonically restrained with a force constant of 1000 kJ.mol<sup>-1</sup>.nm<sup>-2</sup> to equilibrate the solvent. The equilibrated systems were then

simulated for production in NPT ensemble ( $T = 298$  K and  $P = 1.015$  bar) using the Nosé-Hoover thermostat [44][45] ( $\tau_T = 0.4$  ps) and the Parrinello-Rahman barostat ( $\tau_P = 3.0$  ps). The periodic boundary conditions were used and the time step for integration was fixed at 2 fs with the neighbor list updated every 10 fs. The particle-mesh Ewald method [46] was used to evaluate the electrostatic interactions and the P-LINCS algorithm [47] was applied to restrain bond lengths to their equilibration values. We carried out production runs of 90 and 140 ns for the PF and the SA simulations, respectively, with the atomic configurations collected every 2 ps for subsequent analysis.

## **II.3 Test case 2: Simulation of the DI protein with and without LDAO**

### *II.3.1. PI4KA DI ab initio modeling*

In absence of the three-dimensional (3D) structure of PI4KA DI and available homologs we used an *ab-initio* approach to create a molecular model of the protein (see §II-4). For this purpose, we used the I-TASSER [29] and QUARK [30] web servers with the PI4KA DI amino acid sequence as an input. We used the default parameters without specifying additional restraints, secondary structure or excluding some templates to guide the modeling. We chose these two modeling servers, since it was shown previously [48] that they provide the best results for obtaining 3D models and performing atomic-level structure refinements of proteins with unknown 3D structure. These two web servers use different approaches to obtain 3D structure of a given protein. Briefly, I-TASSER uses an iterative atomic-level structure assembly refinement to generate the three-dimensional atomic models from multiple threading alignments and iterative structural assembly by using a replica-exchange Monte Carlo simulation technique. In addition, this method also uses a reduced model of the protein where each residue is represented by its  $C_\alpha$  and side-chain center of mass. Then, the 3D shape of the reduced model is compared with known proteins with similar 3D structure that have the same biological function. On the other hand, QUARK uses a semi-reduced model to represent protein

residues by considering in the computation only the full backbone atoms and the side-chains centers of mass. A neuronal network is used to predict the additional structural features. The global fold is then generated by assembling the small fragments with also replica-exchange Monte Carlo simulations with a composite knowledge-based force field. In this last step, QUARK uses small fragments of 1 to 20 residues and the flexibilities of them whereas I-TASSER uses fragments of 3/9 residues [49]. We finally compared the quality of the two DI models obtained with I-TASSER and QUARK with secondary structure predictions of other predictor servers (such as RaptorX [50], Phyre2 [51], GenThreader [52], PredictProtein [53]). We found that all the programs predicted roughly the same six helices for all DI models but with differences in helix lengths (see Figure S2 in the Supporting Information). For instance I-TASSER, RaptorX and Phyre2 predict a shorter first helix (3 amino-acids of difference) in the protein compared to GenThreader, PredictProtein or QUARK).

### *II.3.2 Simulation of the DI protein with and without LDAO*

We simulated the two DI protein models in presence and absence of LDAO with same simulation parameters as for the pure micelle simulations. The protein and the LDAO were both modeled with the AMBER99SB-ILDN force field, as this force field give accurate results for the LDAO properties (see the results in the next sections). The initial conformations of these simulations were obtained by placing the DI protein in the center of a cubic box (81 Å of side) with 41 randomly placed LDAO, 20000 TIP3P water and 3 Na<sup>+</sup> ions to neutralize the system. The simulations of the DI with and without LDAO were carried out during ~200 ns, respectively, excluding the equilibration periods (~900 ps).



## II.4 Expression and purification of DI constructs

### II.4.1 Cloning of DI into pGEX-6P-1 and pET28(a+) expression vectors

We used the same procedures as detailed in our previous work (Taveneau, Blondeau & Bressanelli, submitted). Briefly, we amplified by PCR DNA sequences corresponding to DI from the full-length PI4KA gene sequence (Gene Bank accession no: AB384703). We then introduced the DI DNA coding for residues 1135-1265 of PI4KA in two different bacterial expression plasmids: pGEX-6P-1, containing an N-terminal Glutathione S-transferase (GST)-tag sequence; and a customized pET28(a+) vector containing an N-terminal 6-histidine tag with a TEV protease cleavage site. As a result, the expressed proteins comprise 230 and 21 extra residues upstream the 131 DI residues for GST-DI and his6-DI, respectively. DI inserts were sequenced prior to expression tests (GATC biotech, Germany).

### II.4.2 Expression of GST-DI and his6-DI

The bacterium strains used for protein expression were *E.Coli* BL21(DE3) Rosetta for his6-DI and *E.Coli* C41(DE3) for GST-DI. After transformation with the recombinant pGEX-6P-1 or pET28(a+), a single colony was sub cultured first into 5mL and then into 500mL of fresh LB medium supplemented with 100 µg/mL ampicillin for PGEX-6P-1 construction, 50 µg/mL kanamycin for pET28(a+) construction and 34 µg/mL chloramphenicol for both subcultures at 37°C under agitation. The third day, this subculture was diluted in 5 L of fresh medium with appropriate antibiotics to a final optical density about 0.1 at a wavelength of 600nm (OD600). Cells were cultivated at 37°C with shaking until they reached an OD600 about 0.8. Protein expression was then induced at 12°C with 0.3mM Isopropyl-β-D-thiogalactopyranoside (IPTG). Cells were then incubated overnight and finally harvested by centrifugation of 20 minutes at 3,500xg. Pellets were stored at -20°C.

#### *II.4.3 Purification of GST-DI and his6-DI*

Purification protocols were substantially the same for the two constructs. The entire process was realized at 4°C and all buffers were supplemented with LDAO. Pellets from a 4 L culture were suspended in 40 mL of lysis buffer implemented with 1 tablet of protease inhibitor cocktail (ref 11836170001, Roche) and about 250 U of benzonase (ref 70746-3, Merck-Millipore). Lysis was performed using sonication 3x2 minutes with pulses of 3 seconds at 30% amplitude. The lysate was clarified by two successive centrifugations: the first one for 12 minutes at 8.000g, and the second one for 30 minutes at 40.000g. The supernatant was then filtered with a 0.45 µm PES membrane filter. The lysate was charged at 0.5 mL/min in a GStrap FF or Histrap HP column of 1 mL, depending on the fusion tag, connected to an AKTApurifier 10 liquid chromatography system. The column was washed at 1ml/min until OD280nm baseline was reached. Proteins were eluted with elution buffer at a flow rate of 0.5 ml/min and collected in fractions of 1 mL. Buffer compositions are described in Table S2 in the supporting information. The sample was concentrated in 500µl and loaded on a Superdex 200 10/300 GL column (GE Healthcare), pre-equilibrated in 30mM Tris-Hcl pH 7.4, 300mM NaCl, 10% glycerol, 2mM LDAO. The purified protein in the flow-through was concentrated by ultrafiltration (with a cut off of 50kDa to alleviate LDAO micelle concentration) to a final protein concentration of 0.53 mg/mL for GST-DI and with a cut off of 10kDa to a final concentration of 0.1mg/mL for his6-DI. Each purification step was analyzed by 14% tris tricine SDS-PAGE. The samples were then flash frozen and stored at -80°C.

#### *II.4.4 Biochemical and biophysical analyses*

*Gel filtration analysis.* In order to check the importance of detergent in GST-DI solubility, three purifications were carried out following the purification protocol above. Only the amount of LDAO was changed from the wash step of the affinity column to the end of the purification. Three conditions were tested: without LDAO, 1mM LDAO (*i.e.* below the CMC) and 2mM LDAO (*i.e.* slightly above the CMC). 500  $\mu$ L of the more concentrated fraction of the affinity column were injected on Superdex 10/300 GL in order to avoid the intermediate concentration step. Gel filtration chromatograms were superimposed to evaluate the solubility of DI under these conditions.

*SEC-MALLS analysis of GST-DI.* For DI-LDAO stoichiometry determination, SEC-MALLS-RI (refractive index) analysis of GST-DI was carried out using Superdex 200 10/300 GL Increase column (*GE Healthcare*) connected to Shimadzu HPLC system. In addition to absorbance at 280 nm (UV detection), light scattering was detected with a *mini DAWN TREOS* system (Wyatt Technology) and refractive index measured with an Optilab T-rEX differential refractometer (Wyatt Technology). 30  $\mu$ g of protein were injected at a flow rate of 0.5 ml/min on the column equilibrated in 30mM Tris-HCl pH 7.4, 300mM NaCl, 10% glycerol, 2mM LDAO. Protein was collected in 500  $\mu$ L fractions, precipitated with TCA and analyzed by 14% tris-tricine SDS-PAGE. Molar mass of proteins, micelles and protein-detergent complexes were determined from the triple detection data (MALLS, UV and refractive index) with ASTRA 6.1 software (Wyatt Technology) using protein conjugate analysis with a refractive index increment ( $dn/dc$ )  $0.1804 \text{ ml.g}^{-1}$  predicted thanks to SEDFIT software calculation module [54]. A control experiment was performed by injecting 48  $\mu$ g of GST alone in the same

buffer as GST-DI. 200  $\mu$ g of Bovine Serum Albumin was used to calibrate the detector constant.

*Tryptophan fluorescence analysis.* Degree of tryptophan burial was evaluated for his6-DI by measuring tryptophan fluorescence on a PerkinElmer LS50B fluorimeter at 18°C. Excitation was performed at 295 nm and emission was measured between 305 and 400 nm (both with slit width 2.5nm) at a speed of 120nm/min.

### **III. Results and discussions**

#### **III.1 Simulation of pure LDAO micelle**

In the next six sub-sections, we will focus on the validation of our LDAO RESP by comparing the simulations results with available simulation and experimental data.

##### *III.1.1 Aggregation process of the LDAO monomers into a single micelle*

In Figure 3 we present a few representative snapshots of the aggregation process of the 104 LDAO monomers into a single micelle as a function of simulation time for the Amber12SB simulation. As shown in Figure for this system, it took  $\sim 5$  ns for that all the LDAO monomers form small clusters and additional  $\sim 35$  ns for these small clusters aggregate into a single micelle. To compare the LDAO aggregation process for the three force fields, we report the number of LDAO cluster,  $N(t)$ , as a function of time (Figure 3 panel b). As previously done for pure micelles with DPC [55] and  $\beta$ -dodecyl maltoside (DDM) surfactants [56], we consider that two LDAO molecules belong to the same cluster if any of their alkyl chain carbon atoms are within 4.1 Å of each other. As shown in Figure 3(b), the aggregation kinetics differs significantly amongst the three force fields (see below). Indeed, the LDAO aggregation process is found unexpectedly,  $\sim 2$  times faster in the Amber99SB-ILDN and Amber12SB ( $\sim 30 - 40$  ns) runs than in the GAFFLipid simulations where it took approximately  $\sim 60$  ns for all the LDAO monomers to form a stable micelle (see Figure S3 in the supporting information). It is

not clear to us why we obtain a slower aggregation kinetic in case of the GAFF simulation, however, as we will see in the next section the micelle properties obtained in the GAFFLipid are very close to those obtained in the Amber99SB-ILDN and Amber12SB simulations than the other ones. The aggregation kinetics obtained in the and Amber99SB and Amber12SB runs are similar to the that obtained by Lorenz et al.[15] for the OPLS-AA force field. As previously observed for other micellar systems (see for instance [20] and reference therein), in our results micelle aggregation process occurs with two different time scales. To characterize these time scales [20], we fit the  $N_{\text{cluster}}(t)$  with a double exponential function:  $c_1 e^{-t/\tau_1} + c_2 e^{-t/\tau_2}$  where  $N_{\text{cluster}}(t_{\infty}) = 1$ . In this equation,  $\tau_1$  and  $\tau_2$  correspond to the relaxation times for a faster and for a slower aggregation process, respectively [57]. As for the exponential parameters, if one hand the sum of  $c_1$  and  $c_2$  corresponds approximatively to the initial number of monomers, given the separation of time scales between  $\tau_1$  and  $\tau_2$ ,  $c_2$  can be identified as the initial number of clusters for the slower self-aggregation kinetics. The  $\tau_1$ ,  $\tau_2$ ,  $c_1$  and  $c_2$  values computed from the three self-assembled simulations are reported in Table 1. We first notice that the faster kinetics has a time scale  $\tau_1 < 1$  ns for all the three force fields. This faster aggregation process is followed for Amber99SB-ILDN and Amber12SB by a relaxation time  $\tau_2 = 5 - 6.2$  smaller than the  $\tau_2$  value (24.9 ns) computed in the GAFFLipid simulation. For sake of comparison, the  $\tau_1$  and  $\tau_2$  values obtained in the Amber99SB-ILDN and the Amber12SB simulations are found similar to the values obtained previously [58] for pure  $\beta$ -dodecyl maltoside (DDM) micelle with 132 monomers (0.6 and 3.5 ns, respectively) modeled with the GLYCAM force field [56].

### III.1.2 Size and shape of the micelle

The sizes of the micelles in the six simulations were compared by computing their radii of gyration,  $R_g$ , and their effective radius,  $R_M$ . Figure 4 depicts the time evolution radius of gyration for the 3 SA simulations. We first remark that the  $R_g$  values are stable from 40 ns till the end for the Amber99SB-ILDN and Amber12SB simulations with an average value around 18.5 Å. In case of the GAFFLipid simulation a stable  $R_g$  value is obtained later (after ~55 ns), in line with the slower timescale of the aggregation process and the presence of LDAO clusters in box (Figure S3). The  $R_g$  values obtained from the SA simulations are the same than those obtained in the PF simulations (Table 2). Also, the micelle sizes obtained in the six simulations are in good agreement with previous simulations of LDAO micelle modeled with the CHARMM22 (19.2 Å) [12] and OPLS-AA (17.6 Å) [15] force fields. To compare our micelle sizes with those obtained from the SANS experiments (see by Barlow et al. [11] and Sardet et al. [59]) and from previous simulations [15], we computed the micelle mean effective radius,  $R_M$ , for a solid sphere with uniform density [60] using the following expression:

$$R_M = \sqrt{\frac{5}{3}} \langle R_g \rangle \quad (1)$$

The micelle mean  $R_M$  values are close to 24.0 Å and agree well with SANS experiments (22 - 24.4 Å) [11][59] and from the simulation with the OPLS-AA force field (22.7 Å) [15].

In Figure 5, we present a pictorial view of a typical final snapshot of a self-assembled micelle at the end of the Amber12SB simulation. As shown in Figure 5, the micelle presents pronounced ellipsoidal shape. In order to obtain the micelle shape parameters, we compute the three ellipsoid semi-axis,  $a$ ,  $b$ , and  $c$ , of the equivalent ellipsoid from the principal moment of inertia (MOI) [61]:

$$\begin{aligned} I_1 &= \frac{M}{5} (b^2 + c^2) \\ I_2 &= \frac{M}{5} (a^2 + c^2) \\ I_3 &= \frac{M}{5} (a^2 + b^2) \end{aligned} \quad (2)$$



Where  $M$  is the total mass of the micelle and  $I_1$ ,  $I_2$  and  $I_3$  (with  $I_3 > I_2 > I_1$ ) are the micelle MOI. The mean ratio  $I_1/I_3$  for the three force fields is in range of 1.43 - 1.49 close to the experimental estimate (1.45) [6][7]. In addition to the  $I_1/I_3$  ratio, we have also computed the asymmetry parameter,  $\alpha$ , [62] a SAXS/SANS metric used to evaluate the deviation of the shape of an aggregate from the sphericity, defined as :

$$\alpha = \frac{(2I_1 - I_2 - I_3)}{(I_1 + I_2 + I_3)} \quad (3)$$

An elliptical micellar shape corresponds to an  $\alpha$  larger than 0 [62][63]. The  $\alpha$  parameters reported in Table 2 for the three simulations are very close to each other, do not significantly depend on the protocol used to construct the micelle, and their average value is near  $\sim 0.17$ . Remarkably, the computed  $\alpha$ 's are very similar to the experimental values ( $\sim 0.20$ ) obtained from SANS experiments [11][15][13]. Finally, we notice that the average computed micelles semi-axis ( $a = 20.0 \pm 0.8$  Å,  $b = 24.0 \pm 0.3$  Å and  $c = 29.1 \pm 0.8$  Å) agree well in with those obtained by Marchetti *et al.* [12] from simulations with the CHARMM22 potential ( $a = 20.4 \pm 0.3$ ,  $b = 23.9 \pm 0.2$  and  $c = 26.6 \pm 0.6$  Å) and by SANS experiments ( $a = 19.4 \pm 0.2$ ,  $b = 19.4 \pm 0.2$ ,  $c = 30.5 \pm 0.5$  Å) [7].

### III.1.3 Micelle microstructure

To investigate the spatial distribution of the components of the micelle system, we computed their mean radial mass density profile  $\rho(r)$  as function of the distance  $r$  from the micelle center of mass (COM,  $r = 0$  Å), see Figure 6. Here, the  $\rho(r)$  functions for the solvation water, the LDAO molecules, their heads and tails were computed every 2 ps over the last 50 ns of each simulation. We find that the density profiles of these species are very similar for the six simulations, with the exception of the alkyl chain  $\rho(r)$ 's where some differences exist, in particular near the micelle COM. As we will see below, these distinctions are likely to be related to conformational dissimilarities near the end of the LDAO hydrophobic chain. The computed  $\rho(r)$ 's also show that the alkyl chains extend from 0 to  $\sim 20$  Å and reach a maximum

density of  $0.77 \text{ g.cm}^{-3}$  near the center of the micelle, which is close to the dodecane bulk density ( $0.74 - 0.79 \text{ g.cm}^{-3}$ ) [11][64]. We also find that the LDAO headgroups extend from 16 to  $\sim 26 \text{ \AA}$  in all simulations it corresponding  $\rho(r)$  has a strong peak near  $\sim 22 \text{ \AA}$ , which corresponds, approximatively, to the value of  $R_M$  discussed in the previous section. A similar result ( $21.5 \text{ \AA}$ ) was obtained by Lorentz et al. [15] As far as water is concerned, its corresponding  $\rho(r)$ 's in Figure 6 shows that the solvent penetrates deeply into the micelle headgroup and share some contacts with the micelle hydrophobic surface (see next section).

### III.1.4 Volume and Interfacial properties of the micelle

By neutron reflectometry and SANS experiments, Barlow et al. [11] and Kakitani et al. [6] estimated the molecular volume of a LDAO monomer inside the micelle, or  $V_{LDAO}^V$ , within  $410.2 - 432.9 \text{ \AA}^3$ . In order to compare this value with those obtained from the 6 simulations, we used the *trjVoronoi* program [65] that uses some of the voro++ classes and libraries [66] to compute, at each simulation time frame, the Voronoi volume of each atom of the system using the Voronoi-Delauney tessellation [67]. As previously [20], explicit hydrogen atoms were excluded in the calculation. Thus, we obtained an average  $V_{LDAO}^V$  value for the six simulations of  $430.8 \pm 4.0 \text{ \AA}^3$ , which agrees well with the experimental LDAO volume.

By using the *trjVoronoi* program, one can also compute the solvent accessible surface area (SASA) of the whole LDAO molecule ( $SASA_{LDAO}$ ) and the headgroup ( $SASA_{Head}$ ) (Table 3). The SASA was calculated by summing up the surfaces of the Voronoi polyhedral facets shared between the LDAO atoms and water. The corresponding SASA values reported in Table 3 indicate that the micelles of the six simulations present a similar surface area in contact with water, in particular the computed  $SASA_{LDAO}$  is found within  $127.8 \text{ \AA}^2$  and  $131.5 \text{ \AA}^2$ . The average surface per headgroup ( $SASA_{Headgroup}$ ) is between  $75.9 - 77 \pm 1 \text{ \AA}^2$  and is in agreement with experimental data ( $70 \pm 10 \text{ \AA}^2$ ) [11]. In the last column of Table 3, we also reported the average fraction of surface,  $f_{tail}$ , shared between the alkyl chains and the solvent (with  $f_{tail} =$

SASA<sub>Headgroup</sub>/SASA<sub>LDAO</sub>). The  $f_{tail}$  values (~41%) are found similar between the six micelles and confirm the LDAO headgroup are largely solvated by water.

### *III.1.5. Micelle hydration properties.*

The micelle and headgroup hydration was characterized by computing the number of hydration waters per LDAO molecule as a function of the force field and simulation protocol. To do this, following Lorenz et al.[15], we counted the number of water-LDAO contacts within a distance of 3.5 Å of the micelle surface. Averaging out results from all simulations, we obtained a mean value of  $8.9 \pm 0.2$  water molecules per LDAO in agreement with SANS experiment (8 - 12) [11][6] and previous simulations of Lorentz et al. (~8 - 10) [15].

To investigate further the hydration of the LDAO headgroup, we have also computed the mean radial density function between the water oxygen atom,  $O_w$ , and the oxygen atom of N-oxide group,  $O_N$ , or  $g_{(O_N-O_w)}$ , (see Figure S4 in the Supporting Information). Since in all simulations the N and O atoms in LDAO and the water are represented by the same potential, it is not surprising that the shape of all the  $g_{(O_N-O_w)}$  functions are similar, with a strong peak at ~2.6 Å and a first minimum ~3.4 Å. The only distinguishing feature amongst the  $g_{(O_N-O_w)}$  's is in the first peak width, which is related to hydration of the N-O group. The number of waters in the first shell of the N-O group was found near 2.6, estimated by integrating  $g_{(O_N-O_w)}$  to its first minimum at ~3.4 Å.

Finally, we have investigated the hydrogen bonds formed between the N-oxide and water. We counted a hydrogen bonds (HB) only if the distance  $O_w-O_N$  is within a 3.5 Å cutoff and the  $O_w-H_w-O_N$  angle is greater than 150°[15]. Thus, for the Amber99SB, Amber12SB and GAFFLipid PF/SA simulations we find similar mean HB values of  $2.6 \pm 0.1$ ,  $2.5 \pm 0.1$  and  $2.5 \pm 0.1$ , respectively. These values are slightly higher than those obtained by Lorentz et al. (2.1) [15].

### III.1.6 Conformational properties of LDAO molecules

Differences in the alkyl chain  $\rho(r)$  functions discussed in a previous section suggest that distinct tail conformations might be present in our six simulations. To examine this aspect in more details, we have first computed the LDAO alkyl chain end-to-end distance,  $d_{\text{endToEnd}}$ , between the first ( $C_3$ ) and the last carbon ( $C_{14}$ ). We find that for all simulations  $d_{\text{endToEnd}}$ 's are in range of 11.5 – 12.0 Å, which is smaller than the length of an extended C12 alkyl chain (13.8 Å) [68], but close to the values obtained in previous MD simulations (~13 Å) [19][20][68][69]. This finding confirms that in the micelle hydrophobic core the LDAO alkyl chains are partially folded.

The conformations of the alkyl chain for each simulation were also examined by computing the mean normalized dihedral angle distribution of CCCC dihedral,  $P(\Phi)$ , from the last 60 ns of each simulation. From these functions, we computed the percentage of trans conformation, %trans, for which the angle  $\Phi$  is in an interval of 120 and 240°. The 9 CCCC dihedrals named with a number between 0 and 9 correspond to the sequence of all dihedrals between dihedral  $C_{14}C_{13}C_{12}C_{11}$  and  $C_6C_5C_4C_3$  (see Figure 1). As shown in Figure 7, the percentage of trans conformation for each CCCC dihedrals are very similar between the SA/PF micelles but vary with the force field with higher %trans values in the Amber99SB-ILDN simulation followed by the GAFFLipid and AMBER12SB runs. The mean %trans values for the overall CCCC dihedrals are found to be ~75.0 %, 71.0 %, and 66.4 % for the Amber99SB-ILDN, GAFFLipid and Amber12SB and simulations respectively. The Amber99SB-ILDN and GAFFLipid values are close to those obtained for the CHARMM force field by Marchetti *et al.* [12] for LDAO (~70 %) micelle simulated with the CHARM22 parameters and the experimental value (72%) obtained for SDS from Fourier transform infrared spectroscopy experiment [70].

Following Palazzesi et al. [68], we have examined the orientation of the LDAO molecule inside the micelle by computing the angle ( $\theta$ ) formed by the vector that joins  $C_3$  to  $C_{14}$  atoms and that formed by the  $C_3$  atom with the micelle center of mass (Figure 8D). The distribution

profiles of the angle ( $\theta$ ) for the Amber99SB-ILDN, Amber12SB and GAFFLipid are shown in Figure 8(A-C). Our results indicate that the alkyl chains in the different simulations form an angle around  $\sim 30^\circ - 70^\circ$  with the micelle radial vector in agreement with previous simulation results[68]. The angle distributions in Figure 8 also indicate some differences depending on the simulation protocol. In particular, for the AMBER99SB-ILDN the  $\theta$  angle distribution is wider compared to the other ones.

Finally, we have investigated the mobility of each carbon atom in the LDAO alkyl chain of each micelle as a function of the carbon position (Figure 9), by computing the translational diffusion coefficient ( $D_c$ ) for each carbon of the LDAO alkyl chain. All the curves in Fig. 9 have a similar “U” shape where the higher  $D_c$  values are obtained for the first (C3) and last (C14) carbons of the alkyl chain indicating that these carbons are the most mobile of the micelle core. A similar trend was also obtained by Palazzesi et al. [68] and Marchetti [12] from MD simulations. Our average  $D_c$  values are in the range of  $0.65 - 1.5 \cdot 10^{-5} \pm 0.6 \text{ cm}^2.\text{s}^{-1}$ , which are not far from the  $D_c$  values obtained from Quasi Elastic Neutron Scattering QENS experiments ( $0.1$  to  $1.5 \cdot 10^{-5} \text{ cm}^2.\text{s}^{-1}$ ) [71]. Figure 9 also shows that protocols on the simulation initial conditions have no effect on diffusion coefficients, whereas they depend on the force field. Indeed, we find that the GAFFLipid parameters give results in better agreement with experiments than those obtained from Amber99SB-ILDN or Amber12SB. Nevertheless, we point out that the statistical errors on  $D_c$  are in the order of  $0.2 \cdot 10^{-5} \text{ cm}^2 \text{ s}^{-1}$ , thus larger than differences in  $D_c$  amongst force fields.

### *III.2.1 DI – LDAO interactions*

After the validation of our LDAO model against experiments and simulations, we have used our potentials to simulate the interaction between an LDAO micelle and the DI protein, a fragment of the human lipid kinase protein PI4KA. The DI protein being a membrane

interacting segment of PI4KA (DI is a 131-residue protein comprising residues 1135 - 1265 of human PI4KA)

Since the atomic structure of DI is not available, we resorted to *ab-initio* protein modeling for generating atomic models for our simulations. Fortunately, secondary structure prediction algorithms reliably catalogue DI as all-helical protein [28], well in agreement with preliminary UV circular dichroism measurements (Taveneau & Bressanelli, unpublished). This result allows performing *ab-initio* modeling even if the number amino-acids in the DI is greater than 100. Figures S2 and 10 suggest an higher  $\alpha$ -helix content between the I-TASSER and QUARK models in agreement with the values computed with the DSSP program [72][73], 73 and 83 %, respectively. So to scrutinize the dependency of our simulations on the protein model, we carried out self-assembled simulations with two protein models obtained from two independent *ab-initio* prediction algorithms (see Figure 10).

Experimentally, the DI protein can be purified in a non-aggregated form only in the presence of LDAO detergent above its CMC (Figure 11A). We used a combination of multi-angle light scattering, refractive index detection and UV detection coupled to size exclusion chromatography (Figure 11B) to determine the mass of LDAO actually associated to DI. The size exclusion peaks of excess LDAO micelles and LDAO-associated his6-DI proved too poorly separated to do this (data not shown). We therefore used GST-DI and obtain a protein molecular mass of 90.0 kDa and 18.9 kDa of associated detergent for the main peak (Figure 11B). A control with GST only yields a protein molecular mass of 51.1 kDa with no associated detergent. GST is known to form dimers and these values are in good agreement with the molecular masses computed from the protein sequences of dimeric GST-DI (83.7 kDa) and dimeric GST (53.2 kDa). Thus 18.9 kDa of LDAO are associated with 2 DI molecules, or 41 LDAO molecules per DI molecule that justify the use of 41 LDAO in the simulations of the LDAO-protein complex. In this paper, we only discuss results from simulations on the complex formed by the DI protein (I-TASSER and QUARK model) and 41 LDAO modeled by the



Amber99SB-ILDN force field [23], which, as shown in the previous sections, can model the micelle properties sufficiently accurately. It is worth mentioning that simulations with the I-TASSER protein model and the GAFFLipid force field were also carried out, bringing results in line with those reported here (data not shown).

Figure 12(A-B) depicts final snapshots of the LDAO-DI complex with the protein modeled by I-TASSER and QUARK in presence of the 41 LDAO molecules. As we will see below, the simulations produce a variety of results concerning the structure of the protein and its binding to the detergent molecules. To examine this, we firstly analyzed the stability of our DI-LDAO micelle complexes and the complex and the protein sizes by computing the radius of gyration of the components of the protein-LDAO complex either in a detergent environment or in water alone (Figure 13). The aggregation process between the DI and the LDAO is similar between the two systems and it took ~40 ns for the DI and the LDAO form a stable complex with constant  $R_g$  values around 19 Å. If we interest of the proteins alone, their average sizes are similar to the protein  $R_g$  obtained in pure water around 15 Å. These results indicate that, here, the  $R_g$  is not useful metric to discriminate the structural differences between the two models observed in the two simulations. So to go further, we have examined the stability of the initial conformations of the two protein models by computing the time evolution of the proteins alpha carbon,  $C_\alpha$ , root mean square deviations (RMSD) with and without detergent (Figure 14). The minimized structure of the I-TASSER and QUARK models was used as a reference in the RMSD calculations, In pure water and LDAO, the RMSD values from the last 20 ns of each run obtained for I-TASSER are higher than those obtained for QUARK ~4.5 (3.2 Å) and 5.2 Å (4.0 Å), respectively. It suggests that the conformation of the I-TASSER model is not stabilized by the LDAO in contrast to the QUARK protein model. To understand the origin of the structural changes in two proteins, we computed the time evolution of the secondary structure (SS) of the I-TASSER and QUARK in pure water and in the LDAO complex with the help of the DSSP program [72][73] (see Figure S5). The Figures show that in pure water

and (in LDAO), the secondary structure of the I-TASSER protein model is less stable than QUARK one; leading a loss of 8 and (7 %) of its helix content, respectively, compared to the initial structure (73 %). In case of the QUARK model the loss of the alpha helix content is small (3%) in water as well as, in LDAO. In case of the I-TASSER model, the loss of the SS mainly occurs in the 1<sup>st</sup>, 3<sup>th</sup>, 4<sup>th</sup>, and 6<sup>th</sup>  $\alpha$ -helix whereas in the QUARK model the SS loss occurs only at the beginning of the 2<sup>nd</sup>  $\alpha$ -helix.

### *III.2.2 Interfacial properties of the Protein-LDAO micelle complex*

According to Figure 12, the LDAO do not bind to the same protein region depending on the protein model. In order to examine this, we analyzed the number of unique contacts ( $N_{\text{prot-LDAO}}$ ) between the LDAO and the protein. To obtain these numbers, we counted a contact between the protein and LDAO if the distance,  $d$ , between the alkyl chain atoms of the detergent and the protein is smaller than 4.0 Å [74]. We obtained similar averaged overall contacts between LDAO and the protein and the I-TASSER and QUARK models,  $22.6 \pm 2.6$  and  $23.5 \pm 2.4$ , respectively. Comparing the local covering of the protein helices, Figure S6 shows that the DI 2<sup>nd</sup> and 3<sup>rd</sup> helices are similarly covered by LDAO in the both models. Concerning the end of the 1<sup>st</sup>, 5<sup>th</sup> and 6<sup>th</sup> helix they are mainly covered by the LDAO in QUARK model and it is opposite for 4<sup>th</sup>. These differences of the LDAO coverages may explain the changes of the secondary structure of the protein between the two models observed in the previous section). Figures 10 also suggests that in the two protein models two of the four DI tryptophans (1234 and 1236) lie on either side of the fifth helix (Figure 12, orange helix) and may be accessible to the solvent. In Figure 16B we present the hydropathicity plot of the DI protein computed with the Kyte-Doolittle approach (a default window size of 9 residue) [75] with the help of the ExPASy server [76]. We have highlighted in the Figure the four tryptophan of DI (corresponding to Trp1216, Trp1234, Trp1236 and Trp1256 of PI4KA) with a red triangle. A positive hydropathicity signals hydrophobicity, conversely negative values are found for

hydrophilic residues. The program predicts that 3 tryptophans 1234, 1236 and 1256 of 4 have a hydrophobic character and therefore can be buried in the protein core or in contact with the surfactant. The hydrophobic/hydrophilic character (i.e. buried/covered by LDAO or not) of these tryptophans was investigated by fluorescence spectroscopy with a his6-DI in LDAO buffer. The emission spectrum upon excitation at 295 nm displays a clear maximum at 331 nm (Figure 15) showing that all 4 DI tryptophans are in apolar environments. To link up these results with the simulations, we computed the time evolution of protein SASA with the *trjVoronoi* program and deduced the SASA relative difference between the system without detergents ( $SASA_x^{Water}$ ) and the systems with detergents ( $SASA_x^{LDAO}$ ):

$$\sum_x = \frac{\langle SASA_x^{Water} - SASA_x^{LDAO} \rangle}{\langle SASA_x^{Water} \rangle} \quad (4)$$

Figure 16A shows that  $\Sigma_x$  profile for two simulations as function of the DI protein amino acid sequence. The two I-TASSER and QUARK  $\Sigma_x$  profiles differ mainly in the two regions delimited by the end of the beginning of the 1<sup>st</sup>, the end of 2<sup>nd</sup> and 6<sup>th</sup> alpha helix.

During the QUARK simulation, the last three tryptophans (Trp1234, Trp1236 and Trp1256) that belongs to 5<sup>th</sup> and 6<sup>th</sup> helix are in a hydrophobic environment ( $\Sigma_x > 0$ ) and protected from the solvent, whereas Trp1216 located in the 3<sup>rd</sup> helix is located at the interface. Conversely, for the I-TASSER model, only two tryptophans, Trp1216 and Trp1236, are at the interface, whereas Trp1234 and Trp1256 are in a hydrophilic environment. Finally, we point out that in the two models the 4 tryptophans can at some simulation time be buried inside the protein and at other times be covered by detergent (see Figure S7 and S8). This is the case for the Trp1234 and Trp1256 with the I-TASSER model and for the Trp1234 for the QUARK model where buried 57, 83 and 66 % of time, respectively.

As a final remark, by taking into consideration the RMSD the secondary structure behaviors the tryptophan residues behavior, we can conclude that the QUARK model results are in better agreement with experiments than those obtained from I-TASSER.

## IV. Conclusion

This paper has been devoted to the development and testing a new AMBER compatible force field for the LDAO surfactant. Thus, we have derived original RESP charges for the LDAO surfactant that can be used in MD simulations with the AMBER force fields for proteins. We tested and validated these charges with MD simulations of two systems. In the first system, we simulated a pure micelle in water composed of 104 LDAO molecules modeled by the Amber99SB-ILDN, Amber12SB and GAFFLipid force fields. We tested different initial conditions for these systems by using preformed micelles and self-assembling of random distribution of LDAO molecule. Thus, we show that the micelle characteristics (i.e. size, shape, headgroup hydration surfactant alkyl chain conformation) agree well with experiments and with previous simulation data. In particular, we find that the simulated micelles have an ellipsoid shape with a major and minor axial ratio around 1.4 and an effective size around 22 Å. On average, the number water molecules bound to LDAO is  $\sim 9$  and each LDAO headgroup has  $\sim 2.6$  waters in their hydration shell and form  $\sim 2.5$  hydrogen bonds. In case of the LDAO alkyl chains, we find that they are partially folded, with on average a percentage of trans conformation around 71% and that their axis for an angle of  $\sim 30^\circ - 70^\circ$  with the micelle radial vector.

Further, we examined the structural properties of a complex formed by LDAO, modeled by our potential, and the DI protein, a segment of PI4KA in contact with the cellular membrane. In absence of an experimental 3D structure of this protein two DI protein models were constructed with the I-TASSER and QUARK protein modeling algorithms. These two models have different folds. In particular, the QUARK model has a higher percentage of  $\alpha$ -helix than I-TASSER ( $\sim 73$  and  $83\%$ ). Thus, we carried out two sets of simulations of  $\sim 200$  ns MD simulations with 1 protein in pure water and with 41 LDAO molecules. The protein and the surfactant were modelled with the Amber99SB-ILDN force field and our RESP. The

simulations show that the conformation and the secondary structure of the I-TASSER model is less stable in water and in LDAO micelle than QUARK one.

Fluorescence spectroscopy experiments show that the 4 tryptophans in the DI protein (i.e. Trp1216 (Trp82), Trp1234 (Trp100), Trp1236 (Trp102) and Trp1256 (Trp122)) are in an apolar environment. Our simulations of the QUARK model in a LDAO micelle agree well with this observation, and we find that on average Trp1216 and the group of three tryptophans (Trp1234, Trp1236 and Trp1256) are not accessible to water 48 % and 87 % of the simulation time, respectively. In contrast, for the I-TASSER model where we find that only two tryptophans are protected from the solvent during all the simulation time. These results suggest that the QUARK server may provide a better model for the DI protein. Nevertheless, further experiments are also needed to better characterize the LDAO and DI interactions.

## VI. List of Figures

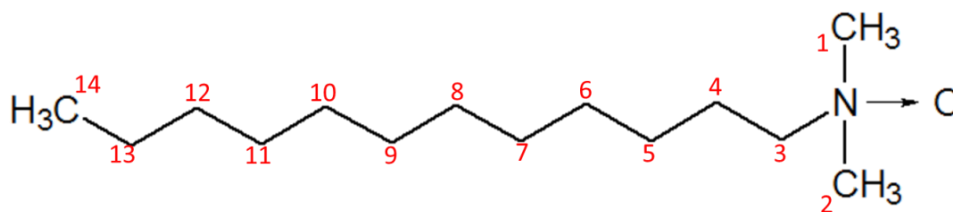


Figure 1: N, N-dimethyldodecylamine-N-oxide (LDAO) molecule with the carbon numeration used for our analysis

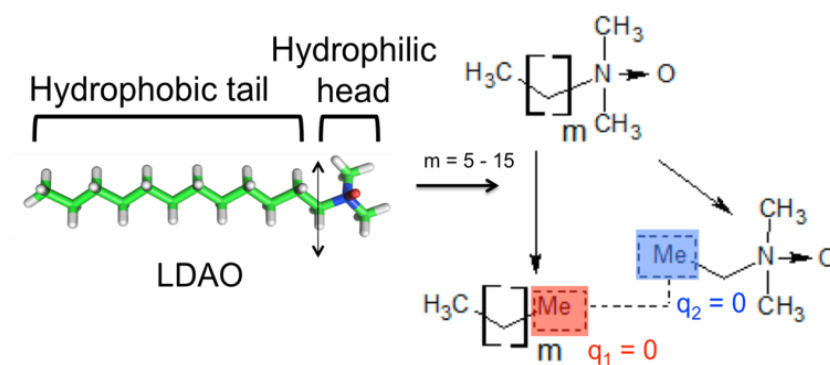
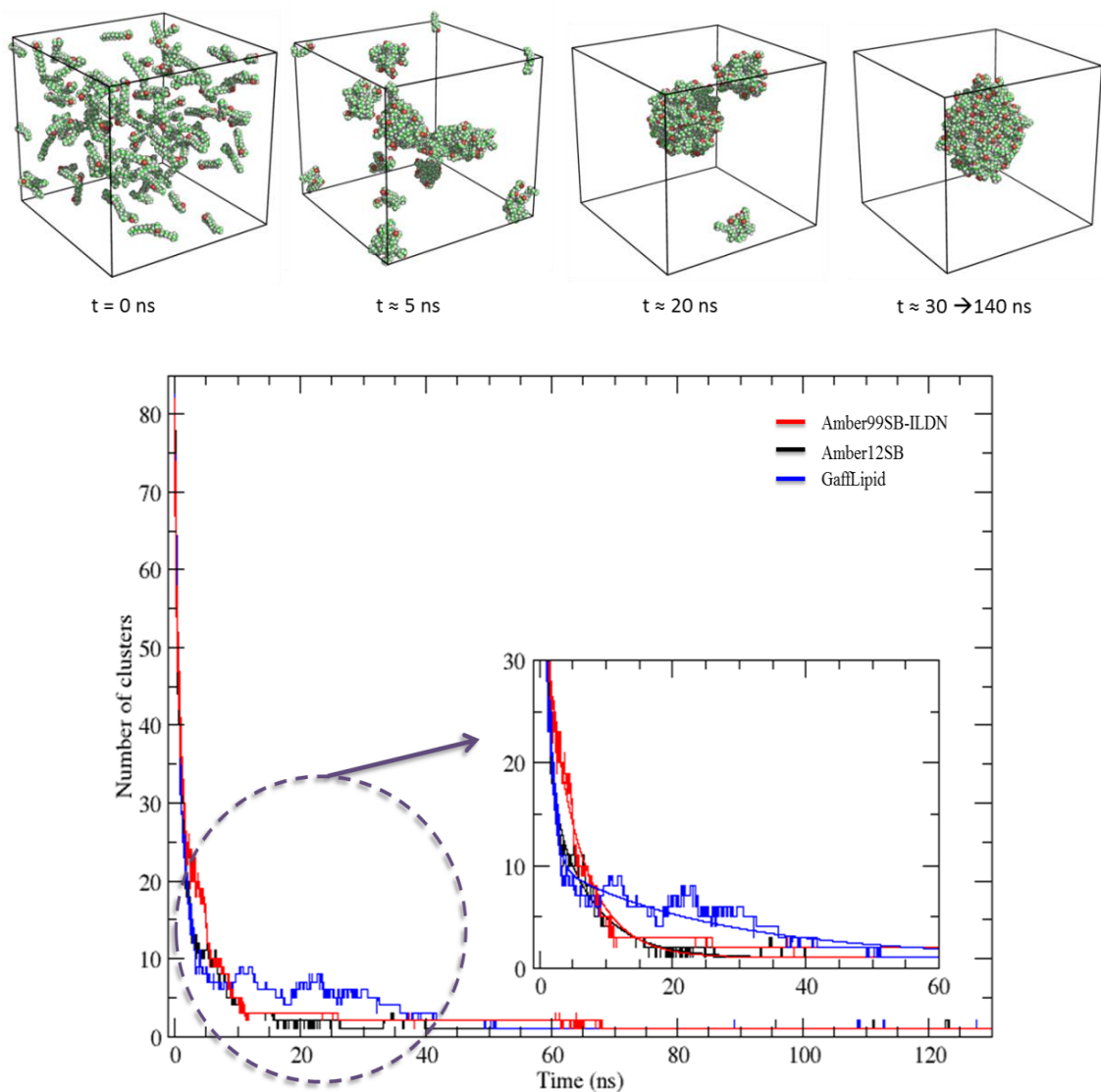


Figure 2: The LDAO molecule on the left and the LDAO molecular fragments for the RESP charges calculations on the right. "Me" stands for the methyl group involved in the Intermolecular charge constraints (see text). With this approach it is possible to obtain a homogeneous library of RESP charges for n-alkyl N, N-dimethylamine-N-oxide with different chain length (here, hexane to hexadecane)





**Figure 3: LDAO self-assembly as a function of simulation time. On the top, the snapshots of the Amber12SB simulation for a self-assembled micelle and on the bottom, the number of LDAO clusters as a function of simulation time.**

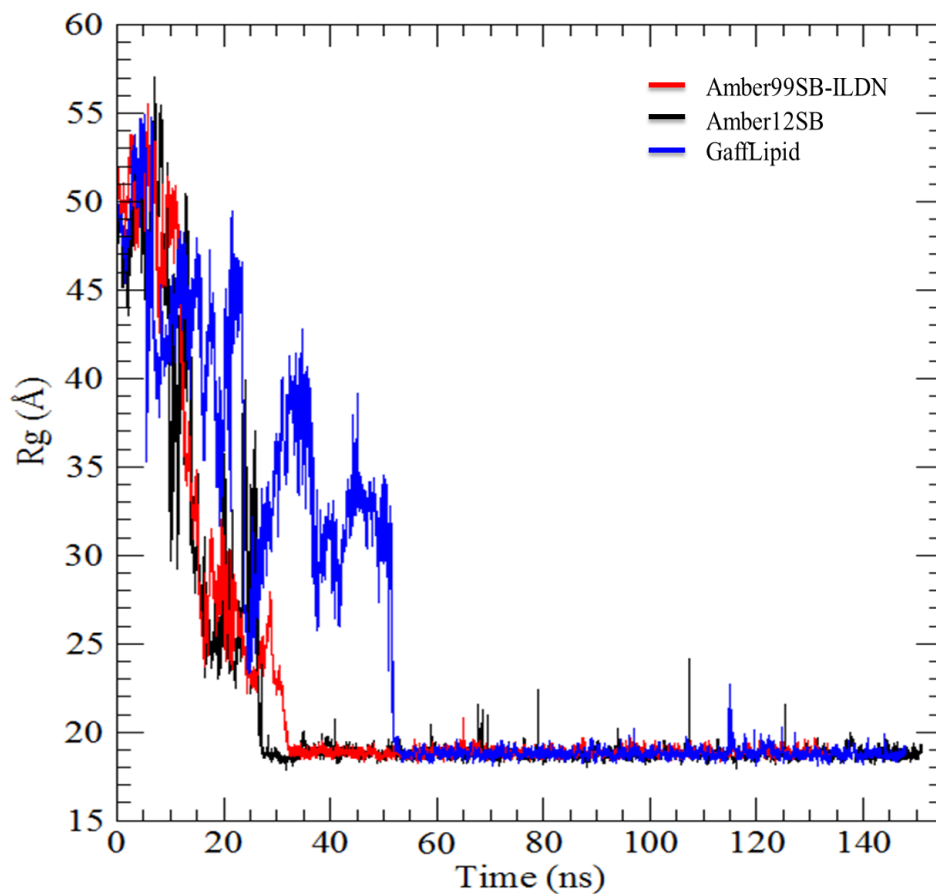


Figure 4: Micelle Radius of gyration as a function of time.

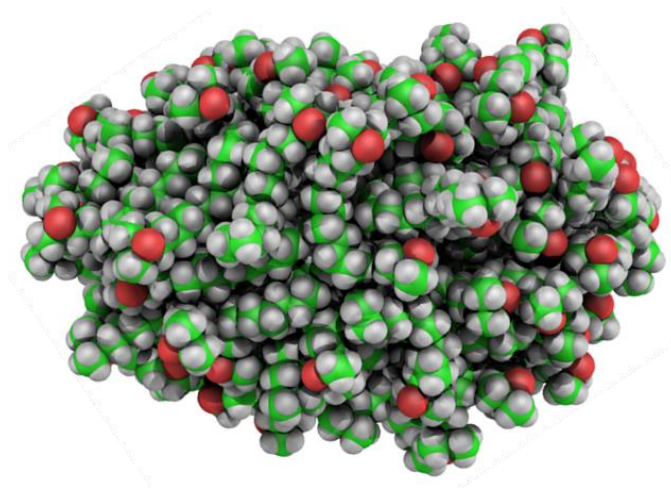


Figure 5: Final snapshot of the self-assembled Amber12SB simulation represented as van-der-Waals spheres. The carbon atoms are in green, the oxygen atoms in red, the hydrogen atoms in white and the nitrogen atom in blue. This figure was produced with the PyMOL program [77].

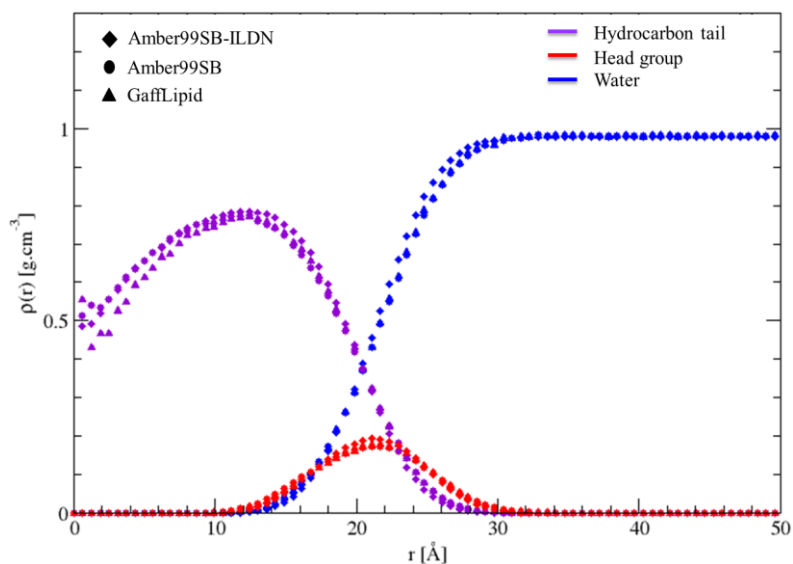


Figure 6: Radial density profile for the Amber99SB-ILDN (♦), Amber12SB (●) and GAFFLipid ( ▲ ) simulations. Note that the values for the PF and SA initial starting condition are superposed. Water, LDAO headgroup and alkyl tail are in blue, red and violet, respectively.

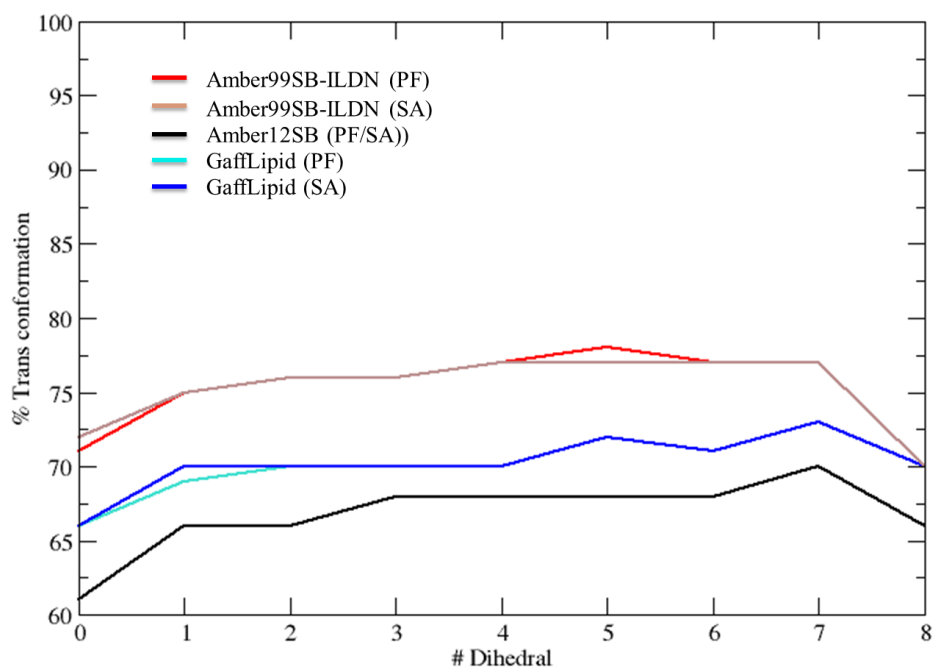


Figure7: The average percentage of *trans* conformation, for each CCCC dihedral angle: 0 is the dihedral at the end of tail and 9 is the dihedral near to N-oxide group (see text).

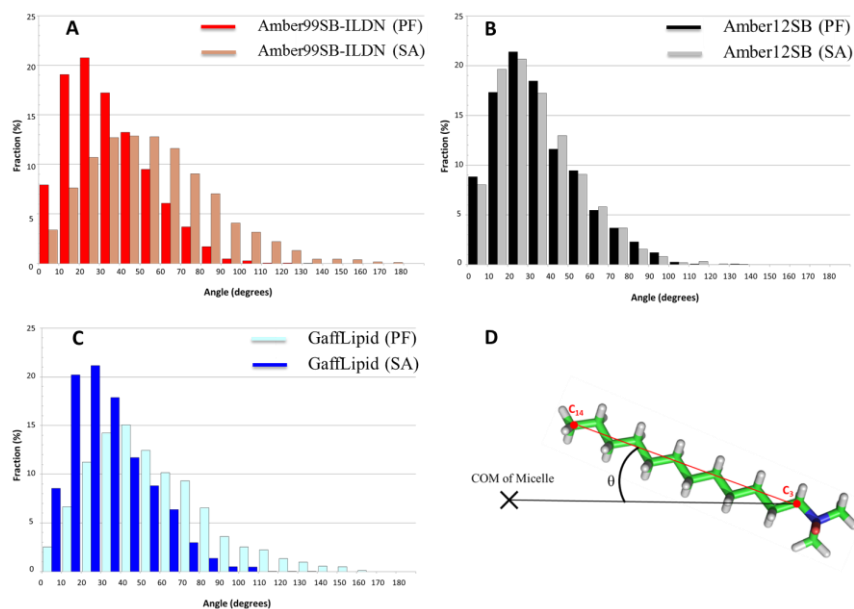


Figure 8: Distribution of the  $\theta$  angle between the LDAO and the micelle COM for the Amber99SB-ILDN (A), Amber12SB (B) and GAFFLipid (C) simulations. The D panel is the scheme for the definition of the  $\theta$  angle [68].

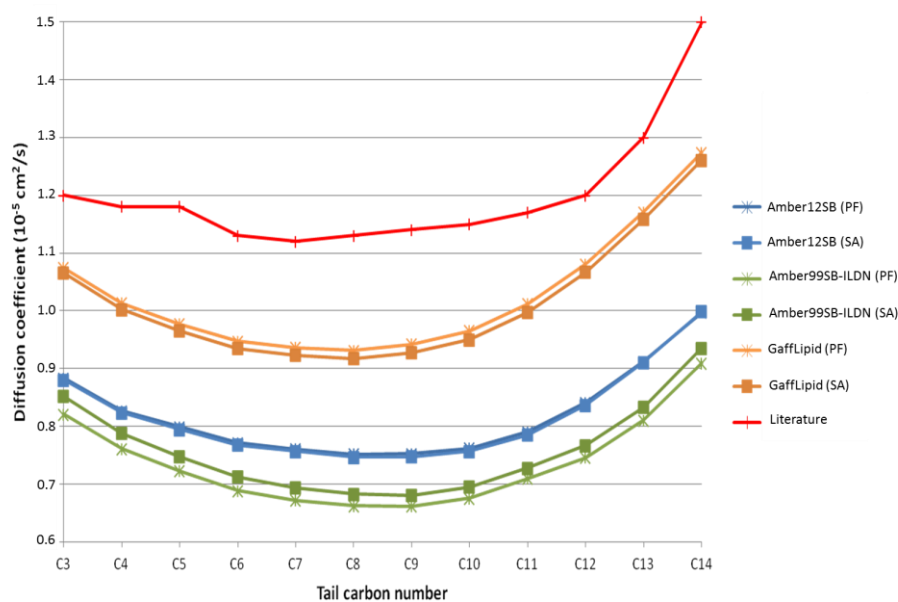
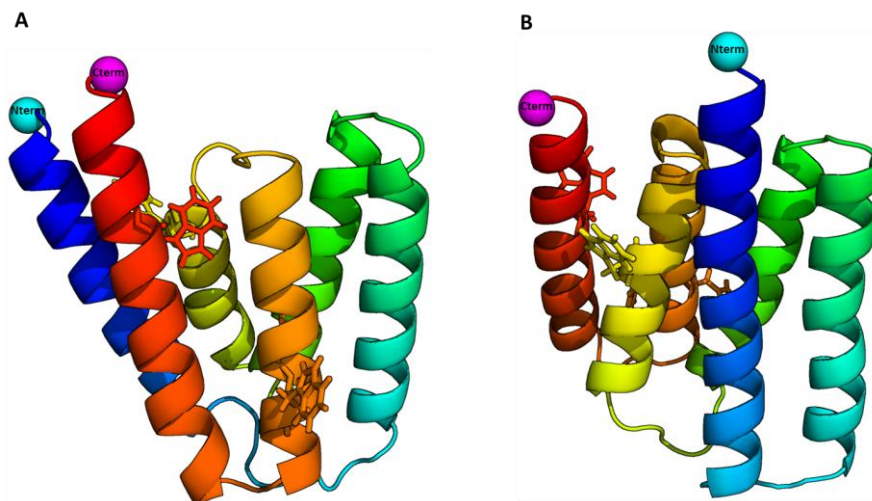
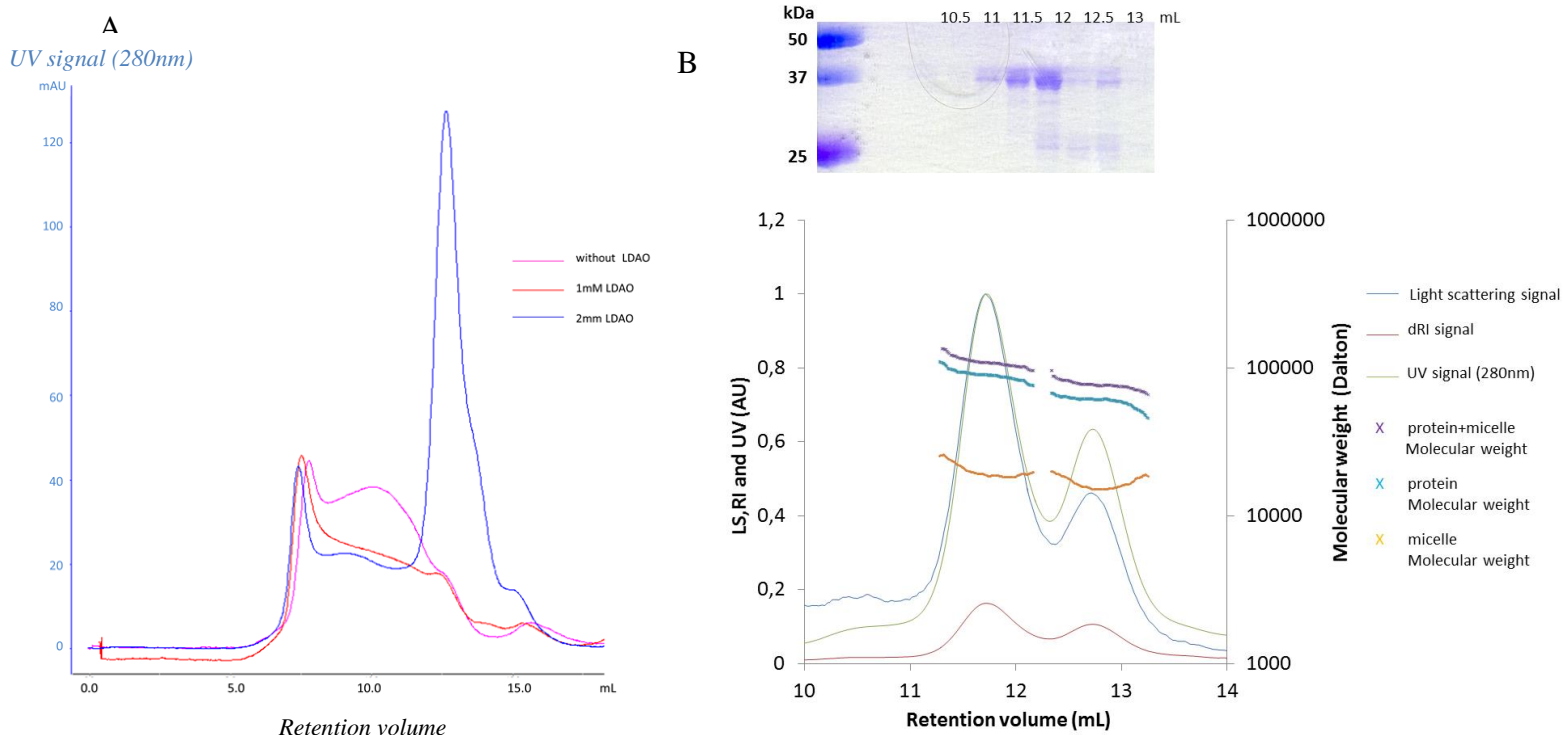


Figure 9: Diffusion coefficient of the carbon atoms of alkyl tail for the Amber12SB, Amber99SB-ILDN and GAFFLipid force fields. The C3 atom is the carbon nearest to the head group and the C14 atom is the carbon nearest to the end of tail. We compared our results with the SDS micelles [68] in red.



**Figure 10:** 3D representations of the DI protein models obtained with the I-TASSER (A) and with the QUARK (B) web servers. The cyan and pink balls show the N and C terminal residues, respectively. To show the differences in topology, we aligned the two proteins in the Figure according to their helices 2 and 3 (in green) and obtained of RMSD of  $\sim 1.7$  Å between the two models. The relative positions of the 4 tryptophans (Trp82, Trp100, Trp102 and Trp122) corresponding to their positions with their position in the (PI4KA) sequence (1216, 1234, 1236 and 1256) are shown in sticks. This Figure was produced with the PyMOL program [77].



**Figure 11:** Experimental characterization of LDAO-DI interaction. In the figure A, the size-exclusion chromatography experiments of GST-DI with different amounts of LDAO. The protein was eluted under three conditions: no detergent (pink), detergent under the CMC (1mM, red) and at the CMC (2mM, blue). Only the experiment with 2mM of LDAO permits the elution of GST-DI in a single pic at 14mL. The experiments were carried out on Superdex 200 10/300 GL column during the purification of DI. In the figure B, determination of the aggregation number of the detergent molecules in the complex GST-DI with LDAO using SEC-MALLS experiment. GST-DI alone was first eluted following by a heterogeneous mix of GST-DI plus DI composing the second pic of the chromatogram. The corresponding polyacrylamide SDS-Page is displayed on the top.

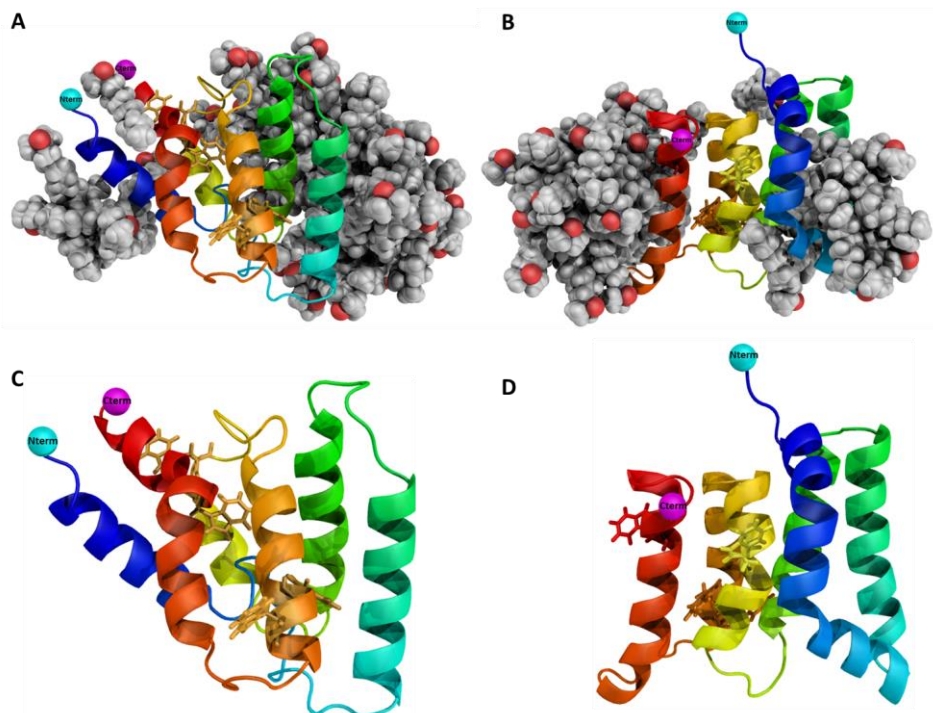


Figure 12: Final snapshot of two simulations, with the I-TASSER model with LDAO (A) and without LDAO (C), and with the QUARK model with LDAO (B) and without LDAO (D). The 4 tryptophans are showed in sticks. In grey, the LDAO molecules and the DI protein is colored blue to red from N- to C-terminus. Please note that to show the differences in topology, we aligned the two proteins according to their helices 2 and 3 (in green) as in Figure 10.

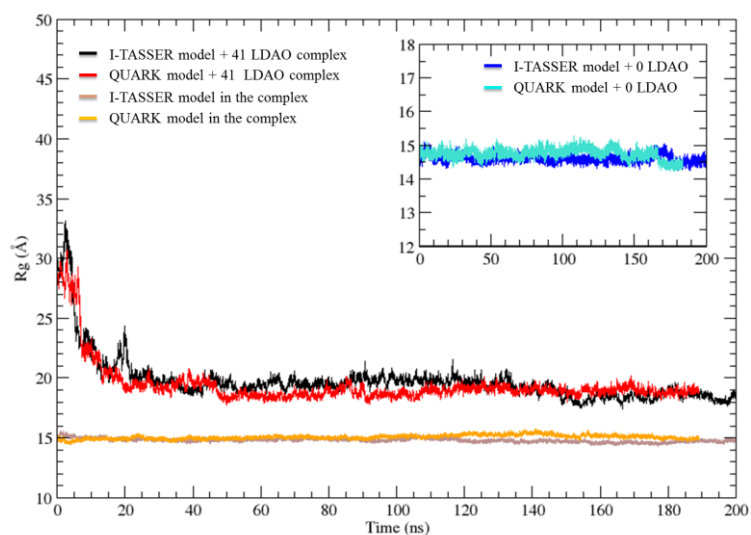


Figure 13: Radius of gyration of the protein-LDAO complex with the I-TASSER model in black and the QUARK model in red. The brown and orange curves represent the  $R_g$  values for the protein (the I-TASSER and QUARK model, respectively) in the complex. The inset corresponds to simulations of the DI protein models in water (the I-TASSER model in blue and the QUARK model in cyan).



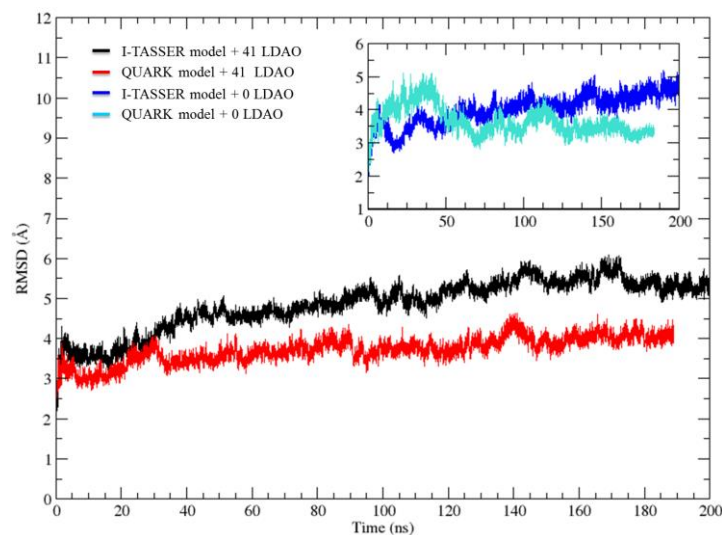


Figure 14: RMSD of protein  $\alpha$ -carbons for the I-TASSER model with 41 LDAO (black) and without LDAO (blue), and the QUARK model with 41 LDAO (red) and without LDAO (turquoise).

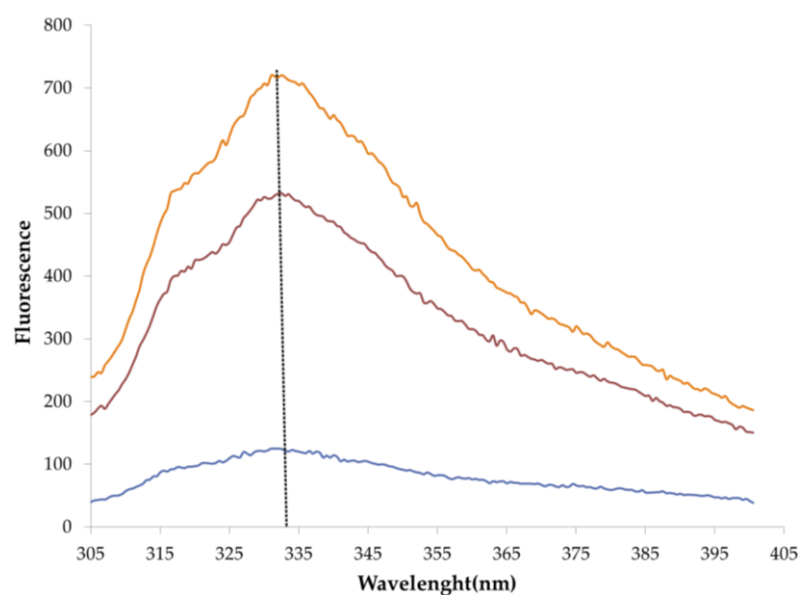
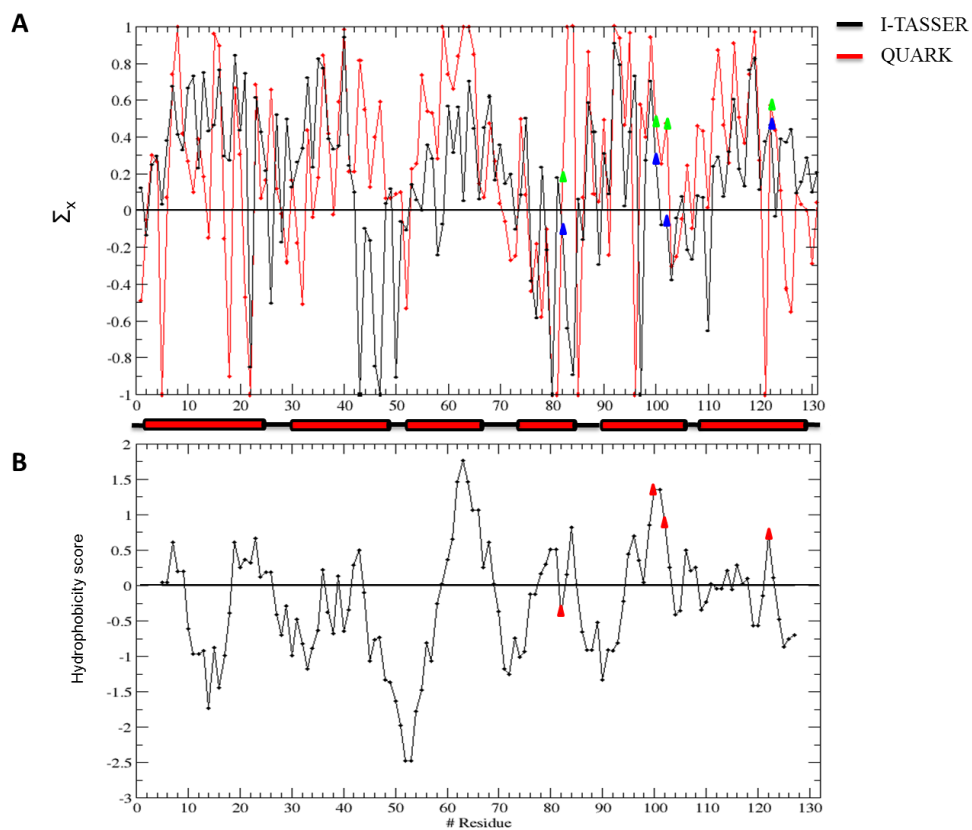


Figure 15: Intrinsic fluorescence emission spectra of his6-DI in LDAO buffer upon excitation at 295 nm. The emission wavelength seems to be independent from the protein concentration. The maximum about 331nm reveals the four tryptophans burial in the micelle or in the hydrophobic core of the protein. Concentrations are 2.7, 1.8 and 0.45  $\mu\text{g/ml}$  for the curves in orange, maroon and blue, respectively.



**Figure 16:** Relative difference in SASA<sub>x</sub> for the two simulations with the I-TASSER (black) and QUARK (red) model in top and hydropathicity plot for the DI protein (bottom). The blue and green triangles represent the tryptophan 1216, 1234, 1236 and 1256 for the I-TASSER and the QUARK model, respectively. The red triangles correspond to these 4 tryptophans in the hydropathicity computed with the Kyte-Doolittle [75] approach used by ExPASy server [76]. For these graphics, positives and negative scores correspond to hydrophobic and hydrophilic characters respectively.

## V. List of Tables

**Table 1: Double-exponential fitting values for the self-assembled simulations (see main text for details).  $\tau_1$  and  $\tau_2$  are in ns.**

System	$c_1$	$\tau_1$ (ns)	$c_2$	$\tau_2$ (ns)
Amber99SB-ILDN	39.3	0.5	35.7	5.0
Amber12SB	58.5	0.7	19.8	6.2
GAFFLipid	70.2	0.9	9.5	24.9

**Table 2: Average dimensions of all the micelles simulated with the pre-assembled (PF) and self-assembled (SA) initial conditions. The maximum error of the radii of gyration in Å is 0.3 Å, the moment of inertia  $I_1$ ,  $I_2$  and  $I_3$  in  $10^4$  a.m.u.nm<sup>2</sup> ( $\pm 0.4$ ), the axial ration  $I_1/I_3$  ( $\pm 0.10$ ) and the asymmetry parameter  $\alpha$  ( $\pm 0.07$ ).**

System		$R_g$	$R_M$	$I_1$	$I_2$	$I_3$	$I_1/I_3$	$\alpha$
Amber99SB-ILDN	PF	18.5	23.9	6.9	5.9	4.8	1.43	0.17
	SA	18.5	23.9	6.0	5.2	4.0	1.49	0.18
Amber12SB	PF	18.7	24.1	6.9	5.9	4.7	1.46	0.18
	SA	18.8	24.3	6.9	6.0	4.7	1.48	0.18
GAFFLipid	PF	18.8	24.3	6.9	6.1	4.7	1.46	0.17
	SA	18.8	24.3	6.9	6.0	4.7	1.46	0.17

**Table 3: Average of LDAO-Water surface area per LDAO molecule for all simulations (see main text for details).**

System			SASA ( $\text{\AA}^2$ )	$f_{\text{tail}}$ (%)
Amber99SB-ILDN	PF	SASA <sub>LDAO</sub>	$131.5 \pm 2.6$	$41 \pm 1.0$
	SA	SASA <sub>Head</sub>	$77.0 \pm 1.0$	
Amber12SB	PF	SASA <sub>LDAO</sub>	$127.7 \pm 3.0$	$41 \pm 1.0$
	SA	SASA <sub>Head</sub>	$75.7 \pm 1.0$	
GAFFLipid	PF	SASA <sub>LDAO</sub>	$130.1 \pm 2.8$	$42 \pm 1.0$
	SA	SASA <sub>Head</sub>	$75.9 \pm 1.0$	

## Supporting Information

Two tables and 6 figures containing the RESP charges and the atom types used to model the LDAO, additional simulation results and list of the chemical compounds in experiments,. This information is available free of charge via the Internet at <http://pubs.acs.org>.

## Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

## **Funding Sources**

This work was supported by the granted access to the HPC resources of CCRT/CINES made by GENCI (Grand Equipement National de Calcul Intensif) under the allocations t2013077082 and t2014077082. We thank also “Maison de la Simulation” and the CEA for the PhD financial support of EK. This work was also supported by ANRS (France REcherche Nord & sud SIDA - hiv Hépatites: FRENSh) through a grant to SB and a predoctoral fellowship to CT and by the French Infrastructure for Integrated Structural Biology (FRISBI) [ANR-10-INSB-05-01].

## **Acknowledgements**

We acknowledge Christophe Velours of the I2BC core biophysical platform for SEC-MALLS analyses. This work has benefited from the facilities and expertise of the Macromolecular Interaction Platform of I2BC (UMR 9198 - <http://www.i2bc-saclay.fr/spip.php?article280>). We also acknowledge Drs Stéphane Roche for advice and assistance in fluorescence spectroscopy and François-Yves Dupradeau (U. of Amiens, France) for helpful discussions during the early stage of this work.

#### IV. References

- [1] V. Kocherbitov and O. Söderman, "Hydration of dimethyldodecylamine-N-oxide: enthalpy and entropy driven processes," *J. Phys. Chem. B*, vol. 110, no. 27, pp. 13649–13655, Jul. 2006.
- [2] V. Kocherbitov, V. Veryazov, and O. Söderman, "Hydration of trimethylamine-N-oxide and of dimethyldodecylamine-N-oxide: An ab initio study," *J. Mol. Struct. THEOCHEM*, vol. 808, no. 1–3, pp. 111–118, Apr. 2007.
- [3] G. G. Privé, "Detergents for the stabilization and crystallization of membrane proteins," *Methods*, vol. 41, no. 4, pp. 388–397, Apr. 2007.
- [4] E. S. Lutton, "Phase behavior of the dimethyldodecyl amine oxide-H<sub>2</sub>O system," *J. Am. Oil Chem. Soc.*, vol. 43, no. 1, pp. 28–30, Jan. 1966.
- [5] K. W. Herrmann, "Non-Ionic-Cationic Micellar properties of Dimethyldodecylamine oxide," *J. Phys. Chem.*, vol. 66, no. 2, pp. 295–300, 1962.
- [6] M. Kakitani, T. Imae, and M. Furusaka, "Investigation of Mixed Micelles of Dodecyldimethylamine Oxide and Sodium Dodecyl Sulfate by SANS: Shape, Size, Charge, and Interaction," *J. Phys. Chem.*, vol. 99, no. 43, pp. 16018–16023, 1995.
- [7] P. Thiagarajan and D. M. Tiede, "Detergent micelle structure and micelle-micelle interactions determined by small-angle neutron scattering under solution conditions used for membrane protein crystallization," *J. Phys. Chem.*, vol. 98, no. 40, pp. 10343–10351, 1994.
- [8] H. Okamura, T. Imae, K. Takagi, Y. Sawaki, and M. Furusaka, "Small-Angle Neutron Scattering Investigation of Supramolecular Assemblies in Ternary Systems of Alkyldimethylamine Oxide/Cinnamic Acid/Water," *J. Colloid Interface Sci.*, vol. 180, no. 1, pp. 98–105, Jun. 1996.
- [9] P. A. Timmins, M. Leonhard, H. U. Weltzien, T. Wacker, and W. Welte, "A physical characterization of some detergents of potential use for membrane protein crystallization," *FEBS Lett.*, vol. 238, no. 2, pp. 361–368, Oct. 1988.



- [10] P. A. Timmins, J. Hauk, T. Wacker, and W. Welte, "The influence of heptane-1,2,3-triol on the size and shape of LDAO micelles Implications for the crystallisation of membrane proteins," *FEBS Lett.*, vol. 280, no. 1, pp. 115–120, Mar. 1991.
- [11] D. J. Barlow, M. J. Lawrence, T. Zuberi, S. Zuberi, and R. K. Heenan, "Small-Angle Neutron-Scattering Studies on the Nature of the Incorporation of Polar Oils into Aggregates of N,N-Dimethyldodecylamine-N-oxide," *Langmuir*, vol. 16, no. 26, pp. 10398–10403, Dec. 2000.
- [12] G. Marchetti, M. Marchi, and M. Le Maire, "Modélisation moléculaire du phénomène du transport du calcium dans la protéine ATPase-Ca<sup>2+</sup> (SERCA1a)," Paris: Université Pierre et Marie Curie., 2006.
- [13] F. Sterpone, G. Marchetti, C. Pierleoni, and M. Marchi, "Molecular Modeling and Simulation of Water near Model Micelles: Diffusion, Rotational Relaxation and Structure at the Hydration Interface," *J. Phys. Chem. B*, vol. 110, no. 23, pp. 11504–11510, Jun. 2006.
- [14] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus, "All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins," *J. Phys. Chem. B*, vol. 102, no. 18, pp. 3586–3616, Apr. 1998.
- [15] C. D. Lorenz, C.-M. Hsieh, C. A. Dreiss, and M. J. Lawrence, "Molecular dynamics simulations of the interfacial and structural properties of dimethyldodecylamine-N-oxide micelles," *Langmuir ACS J. Surf. Colloids*, vol. 27, no. 2, pp. 546–553, Jan. 2011.
- [16] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, "Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids," *J. Am. Chem. Soc.*, vol. 118, no. 45, pp. 11225–11236, Jan. 1996.

- [17] A. Cordero, G. Caltabiano, and L. Pardo, "Membrane Protein Simulations Using AMBER Force Field and Berger Lipid Parameters," *J. Chem. Theory Comput.*, vol. 8, no. 3, pp. 948–958, Mar. 2012.
- [18] A. Kumar, E. Hajjar, P. Ruggerone, and M. Ceccarelli, "Molecular Simulations Reveal the Mechanism and the Determinants for Ampicillin Translocation through OmpF," *J. Phys. Chem. B*, vol. 114, no. 29, pp. 9608–9616, Jul. 2010.
- [19] S. Abel, F.-Y. Dupradeau, E. P. Raman, A. D. MacKerell, and M. Marchi, "Molecular Simulations of Dodecyl- $\beta$ -maltoside Micelles in Water: Influence of the Headgroup Conformation and Force Field Parameters," *J. Phys. Chem. B*, vol. 115, no. 3, pp. 487–499, Jan. 2011.
- [20] S. Abel, F.-Y. Dupradeau, and M. Marchi, "Molecular Dynamics Simulations of a Characteristic DPC Micelle in Water," *J. Chem. Theory Comput.*, vol. 8, no. 11, pp. 4610–4623, Nov. 2012.
- [21] P. Cieplak, W. D. Cornell, C. Bayly, and P. A. Kollman, "Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins," *J. Comput. Chem.*, vol. 16, no. 11, pp. 1357–1377, Nov. 1995.
- [22] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, "Comparison of multiple Amber force fields and development of improved protein backbone parameters," *Proteins*, vol. 65, no. 3, pp. 712–725, Nov. 2006.
- [23] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, and D. E. Shaw, "Improved side-chain torsion potentials for the Amber ff99SB protein force field," *Proteins*, vol. 78, no. 8, pp. 1950–1958, Jun. 2010.
- [24] D. A. Case, T. A. Darden, T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. W. Goetz, I. Kolossv  ry, K. F. Wong, F. Paesani, J. Vanicek, R. M. Wolf, J. Liu,

- X. Wu, A. W. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M. J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P. A. Kollman, “AMBER 12,” 2012.
- [25] C. J. Dickson, L. Rosso, R. M. Betz, R. C. Walker, and I. R. Gould, “GAFFlipid: a General Amber Force Field for the accurate molecular dynamics simulation of phospholipid,” *Soft Matter*, vol. 8, no. 37, pp. 9617–9627, Aug. 2012.
- [26] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, “Development and testing of a general amber force field,” *J. Comput. Chem.*, vol. 25, no. 9, pp. 1157–1174, Jul. 2004.
- [27] T. Balla, “Phosphoinositides: Tiny Lipids With Giant Impact on Cell Regulation,” *Physiol. Rev.*, vol. 93, no. 3, pp. 1019–1137, Jul. 2013.
- [28] C. Harak, D. Radujkovic, C. Taveneau, S. Reiss, R. Klein, S. Bressanelli, and V. Lohmann, “Mapping of Functional Domains of the Lipid Kinase Phosphatidylinositol 4-Kinase Type III Alpha Involved in Enzymatic Activity and Hepatitis C Virus Replication,” *J. Virol.*, vol. 88, no. 17, pp. 9909–9926, Sep. 2014.
- [29] A. Roy, A. Kucukural, and Y. Zhang, “I-TASSER: a unified platform for automated protein structure and function prediction,” *Nat. Protoc.*, vol. 5, no. 4, pp. 725–738, Apr. 2010.
- [30] D. Xu and Y. Zhang, “Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field,” *Proteins*, vol. 80, no. 7, pp. 1715–1735, Jul. 2012.
- [31] F.-Y. Dupradeau, A. Pigache, T. Zaffran, C. Savineau, R. Lelong, N. Grivel, D. Lelong, W. Rosanski, and P. Cieplak, “The R.E.D. Tools: Advances in RESP and ESP charge derivation and force field library building,” *Phys. Chem. Chem. Phys. PCCP*, vol. 12, no. 28, pp. 7821–7839, Jul. 2010.

- [32] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, J. J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, "Gaussian 09, Revision D.01," *Inc., Wallingford CT*, 2009.
- [33] W. J. Hehre, "Ab initio molecular orbital theory," *Acc. Chem. Res.*, vol. 9, no. 11, pp. 399–406, Nov. 1976.
- [34] C. I. Bayly, P. Cieplak, W. Cornell, and P. A. Kollman, "A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model," *J. Phys. Chem.*, vol. 97, no. 40, pp. 10269–10280, Oct. 1993.
- [35] M. L. Connolly, "Solvent-accessible surfaces of proteins and nucleic acids," *Science*, vol. 221, no. 4612, pp. 709–713, Aug. 1983.
- [36] E. Vanqualef, S. Simon, G. Marquant, E. Garcia, G. Klimerak, J. C. Delepine, P. Cieplak, and F.-Y. Dupradeau, "R.E.D. Server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments," *Nucleic Acids Res.*, vol. 39, no. suppl 2, pp. W511–W517, Jul. 2011.
- [37] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," *J. Chem. Phys.*, vol. 79, no. 2, pp. 926–935, Jul. 1983.

- [38] L. Martínez, R. Andrade, E. G. Birgin, and J. M. Martínez, “PACKMOL: a package for building initial configurations for molecular dynamics simulations,” *J. Comput. Chem.*, vol. 30, no. 13, pp. 2157–2164, Oct. 2009.
- [39] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess, and E. Lindahl, “GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit,” *Bioinformatics*, p. btt055, Feb. 2013.
- [40] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans, “Interaction Models for Water in Relation to Protein Hydration,” in *Intermolecular Forces*, B. Pullman, Ed. Springer Netherlands, 1981, pp. 331–342.
- [41] G. Bussi, D. Donadio, and M. Parrinello, “Canonical sampling through velocity rescaling,” *J. Chem. Phys.*, vol. 126, no. 1, p. 014101, Jan. 2007.
- [42] M. Parrinello and A. Rahman, “Polymorphic transitions in single crystals: A new molecular dynamics method,” *J. Appl. Phys.*, vol. 52, no. 12, pp. 7182–7190, Dec. 1981.
- [43] A. Rahman and F. H. Stillinger, “Molecular Dynamics Study of Liquid Water,” *J. Chem. Phys.*, vol. 55, no. 7, pp. 3336–3359, Oct. 1971.
- [44] S. Nosé, “A unified formulation of the constant temperature molecular dynamics methods,” *J. Chem. Phys.*, vol. 81, no. 1, pp. 511–519, Jul. 1984.
- [45] null Hoover, “Canonical dynamics: Equilibrium phase-space distributions,” *Phys. Rev. A*, vol. 31, no. 3, pp. 1695–1697, Mar. 1985.
- [46] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, “A smooth particle mesh Ewald method,” *J. Chem. Phys.*, vol. 103, no. 19, pp. 8577–8593, Nov. 1995.
- [47] B. Hess, “P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation,” *J. Chem. Theory Comput.*, vol. 4, no. 1, pp. 116–122, 2007.

- [48] C.-H. Tai, H. Bai, T. J. Taylor, and B. Lee, "Assessment of template-free modeling in CASP10 and ROLL," *Proteins Struct. Funct. Bioinforma.*, vol. 82, pp. 57–83, Feb. 2014.
- [49] Y. Zhang, "Interplay of I-TASSER and QUARK for template-based and ab initio protein structure prediction in CASP10," *Proteins*, vol. 82 Suppl 2, pp. 175–187, Feb. 2014.
- [50] M. Källberg, H. Wang, S. Wang, J. Peng, Z. Wang, H. Lu, and J. Xu, "Template-based protein structure modeling using the RaptorX web server," *Nat. Protoc.*, vol. 7, no. 8, pp. 1511–1522, Aug. 2012.
- [51] L. A. Kelley and M. J. E. Sternberg, "Protein structure prediction on the Web: a case study using the Phyre server," *Nat. Protoc.*, vol. 4, no. 3, pp. 363–371, 2009.
- [52] D. T. Jones, "GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences," *J. Mol. Biol.*, vol. 287, no. 4, pp. 797–815, Apr. 1999.
- [53] B. Rost, G. Yachdav, and J. Liu, "The PredictProtein server," *Nucleic Acids Res.*, vol. 32, no. Web Server issue, pp. W321–326, Jul. 2004.
- [54] P. Schuck, "Size-Distribution Analysis of Macromolecules by Sedimentation Velocity Ultracentrifugation and Lamm Equation Modeling," *Biophys. J.*, vol. 78, no. 3, pp. 1606–1619, Mar. 2000.
- [55] S. A. Sanders, M. Sammalkorpi, and A. Z. Panagiotopoulos, "Atomistic Simulations of Micellization of Sodium Hexyl, Heptyl, Octyl, and Nonyl Sulfates," *J. Phys. Chem. B*, vol. 116, no. 8, pp. 2430–2437, 2012.
- [56] K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. González-Outeiriño, C. R. Daniels, B. L. Foley, and R. J. Woods, "GLYCAM06: A generalizable biomolecular force field. Carbohydrates," *J. Comput. Chem.*, vol. 29, no. 4, pp. 622–655, Mar. 2008.
- [57] I. A. Nyrkova and A. N. Semenov, "On the Theory of Micellization Kinetics," *Macromol. Theory Simul.*, vol. 14, no. 9, pp. 569–585, Nov. 2005.
- [58] S. Abel, A. Lorieau, B. de Foresta, F.-Y. Dupradeau, and M. Marchi, "Bindings of hMRP1 transmembrane peptides with dodecylphosphocholine and dodecyl- $\beta$ -d-maltoside

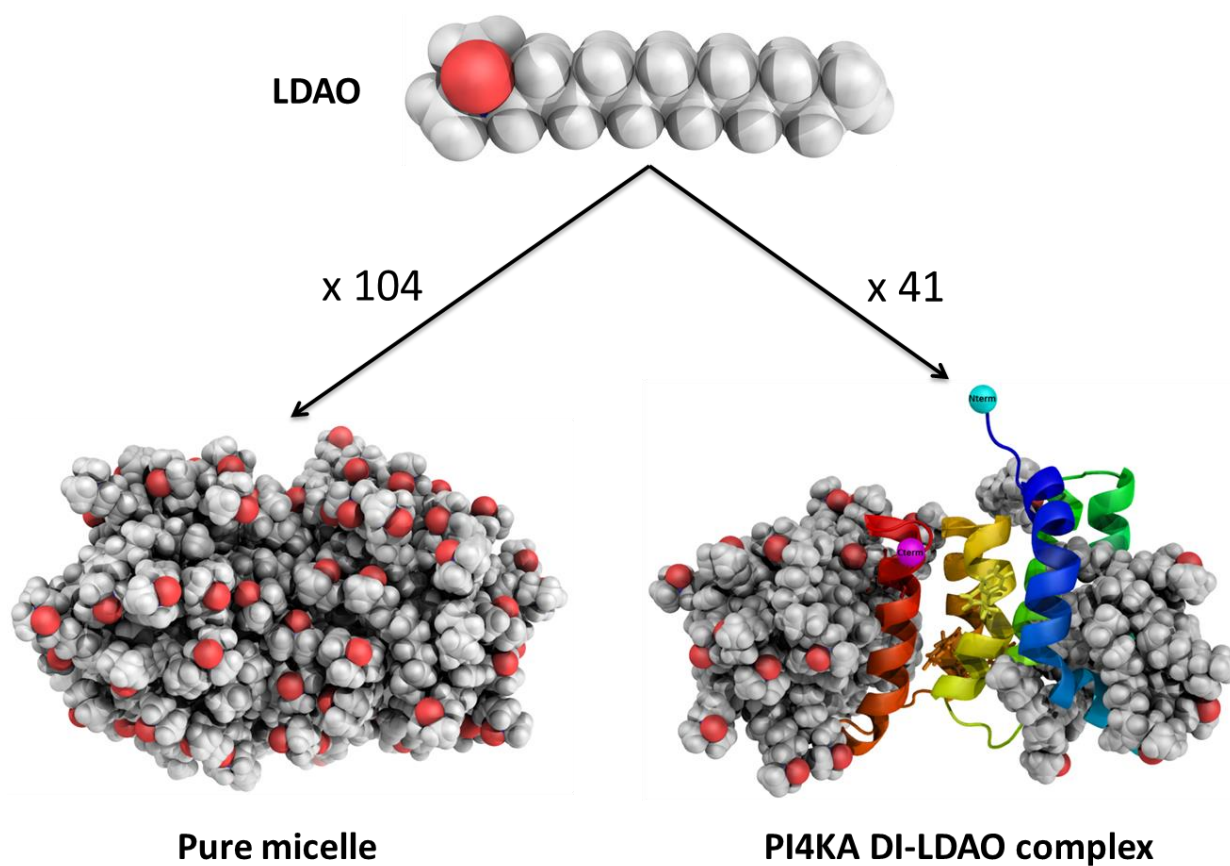


- micelles: A molecular dynamics simulation study,” *Biochim. Biophys. Acta BBA - Biomembr.*, vol. 1838, no. 1, Part B, pp. 493–509, Jan. 2014.
- [59] C. Sardet, A. Tardieu, and V. Luzzati, “Shape and size of bovine rhodopsin: A small-angle X-ray scattering study of a rhodopsin-detergent complex,” *J. Mol. Biol.*, vol. 105, no. 3, pp. 383–407, Aug. 1976.
- [60] R. A. Karjiban, M. Basri, M. B. A. Rahman, and A. B. Salleh, “Molecular Dynamics Simulation of Palmitate Ester Self-Assembly with Diclofenac,” *Int. J. Mol. Sci.*, vol. 13, no. 8, pp. 9572–9583, Jul. 2012.
- [61] S. Abel, F. Sterpone, S. Bandyopadhyay, and M. Marchi, “Molecular Modeling and Simulations of AOT–Water Reverse Micelles in Isooctane: Structural and Dynamic Properties,” *J. Phys. Chem. B*, vol. 108, no. 50, pp. 19458–19466, 2004.
- [62] D. P. Tieleman, D. van der Spoel, and H. J. C. Berendsen, “Molecular Dynamics Simulations of Dodecylphosphocholine Micelles at Three Different Aggregate Sizes: Micellar Structure and Chain Relaxation,” *J. Phys. Chem. B*, vol. 104, no. 27, pp. 6380–6388, 2000.
- [63] L. Huang and P. Somasundaran, “Theoretical Model and Phase Behavior for Binary Surfactant Mixtures,” *Langmuir*, vol. 13, no. 25, pp. 6683–6688, Dec. 1997.
- [64] E. Aicart, G. Tardajos, and M. Diaz Pena, “Isothermal compressibility of cyclohexane + n-hexane, cyclohexane + n-heptane, cyclohexane + n-octane, and cyclohexane + n-nonane,” *J. Chem. Eng. Data*, vol. 25, no. 2, pp. 140–145, 1980.
- [65] M. Marchi, “trjVoronoi is a computational tool written in C++, which uses the voro++ library (version 0.4.3) of C. H. Rycroft to implement the Voronoi tessellation for frames of a GROMACS trajectory. It can compute the Voronoi volume of any given atoms. It can be downloaded from <http://trjvoronoi-cplusplus.googlecode.com/files/trjVoronoi-06-07-2012.tgz>.”

- [66] C. H. Rycroft, "VORO++: a three-dimensional voronoi cell library in C++," *Chaos Woodbury N*, vol. 19, no. 4, p. 041111, Dec. 2009.
- [67] G. Voronoi, "Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Deuxième mémoire. Recherches sur les paralléloèdres primitifs," *J. Reine Angew. Math. - J REINE ANGEW MATH*, vol. 1908, no. 134, pp. 198–287, 1908.
- [68] F. Palazzesi, M. Calvaresi, and F. Zerbetto, "A molecular dynamics investigation of structure and dynamics of SDS and SDBS micelles," *Soft Matter*, vol. 7, no. 19, pp. 9148–9156, Sep. 2011.
- [69] F. Sterpone, G. Briganti, and C. Pierleoni, "Sphere versus Cylinder: The Effect of Packing on the Structure of Nonionic C12E6 Micelles," *Langmuir*, vol. 25, no. 16, pp. 8960–8967, 2009.
- [70] F. Holler and J. B. Callis, "Conformation of the hydrocarbon chains of sodium dodecyl sulfate molecules in micelles: an FTIR study," *J. Phys. Chem.*, vol. 93, no. 5, pp. 2053–2058, Mar. 1989.
- [71] V. K. Sharma, S. Mitra, G. Verma, P. A. Hassan, V. Garcia Sakai, and R. Mukhopadhyay, "Internal Dynamics in SDS Micelles: Neutron Scattering Study," *J. Phys. Chem. B*, vol. 114, no. 51, pp. 17049–17056, 2010.
- [72] R. P. Joosten, T. A. H. te Beek, E. Krieger, M. L. Hekkelman, R. W. W. Hooft, R. Schneider, C. Sander, and G. Vriend, "A series of PDB related databases for everyday needs," *Nucleic Acids Res.*, vol. 39, no. Database issue, pp. D411–D419, Jan. 2011.
- [73] W. Kabsch and C. Sander, "Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features," *Biopolymers*, vol. 22, no. 12, pp. 2577–2637, Dec. 1983.
- [74] R. A. Böckmann and A. Caflisch, "Spontaneous Formation of Detergent Micelles around the Outer Membrane Protein OmpX," *Biophys. J.*, vol. 88, no. 5, pp. 3191–3204, May 2005.

- [75] J. Kyte and R. F. Doolittle, “A simple method for displaying the hydropathic character of a protein,” *J. Mol. Biol.*, vol. 157, no. 1, pp. 105–132, May 1982.
- [76] M. R. Wilkins, E. Gasteiger, A. Bairoch, J. C. Sanchez, K. L. Williams, R. D. Appel, and D. F. Hochstrasser, “Protein identification and analysis tools in the ExPASy server,” *Methods Mol. Biol. Clifton NJ*, vol. 112, pp. 531–552, 1999.
- [77] *The PyMOL Molecular Graphics System, Version 1.7.4*. Schrödinger, LLC.

**Insert Table of Contents Graphic and Synopsis Here**



## SUPPORTING INFORMATION

# A new potential for N,N-Dimethyldodecylamine-N-oxide surfactant based on the Amber force field for Molecular Dynamics Simulations of Micelles and application to a fragment of the lipid kinase PI4KA

*Esra Karakas<sup>1,2,3‡</sup>, Cyntia Taveneau<sup>3‡</sup>, Stéphane Bressanelli<sup>3</sup>, Massimo Marchi<sup>2,3</sup>, Bruno Robert<sup>2,3</sup> and Stéphane Abel<sup>2,3\*</sup>*

<sup>1</sup>Maison de la Simulation, USR 3441, CEA – CNRS – INRIA – Univ. Paris-Sud – Univ. de Versailles, 91191,

<sup>2</sup>Commissariat à l'énergie atomique et aux énergies alternatives, Gif-sur-Yvette Cedex, France DSV/iBiTec-S/SB2SM/LBMS & CNRS UMR 9198

<sup>3</sup>Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Université Paris-Sud, 1 avenue de la terrasse, 91198 Gif-sur-Yvette, France

‡ Equal contributions

\* Corresponding authors: [stephane.abel@cea.fr](mailto:stephane.abel@cea.fr) & [stephane.bressanelli@vms.cnrs-gif.fr](mailto:stephane.bressanelli@vms.cnrs-gif.fr)

## I. List of Tables

**Table S1: The LDAO RESP atomic partial charges for atom type in the Amber99SB-ILDN, Amber12SB and GaffLipid force field.**

Group	Atom type			Charges
	Amber99SB-ILDN	Amber12SB	GaffLipid	
CH <sub>3</sub> Head Group	CT	C8	a3	-0.0960
CH <sub>3</sub> Head Group	HP	HC	hl	0.0790
N-Oxide Head Group	N4	N4	N4	0.1569
N-Oxide Head Group	O	O	O	-0.6041
1 <sup>st</sup> CH <sub>2</sub> Hydrophobic Tail	CT	C8	a3	0.0269
1 <sup>nd</sup> CH <sub>2</sub> Hydrophobic Tail	HC	HC	hl	0.0449
2 <sup>rd</sup> CH <sub>2</sub> Hydrophobic Tail	CT	C8	a3	0.0612
2 <sup>th</sup> CH <sub>2</sub> Hydrophobic Tail	HC	HC	hl	-0.0073
3 <sup>th</sup> CH <sub>2</sub> Hydrophobic Tail	CT	C8	a3	0.0167
3 <sup>th</sup> CH <sub>2</sub> Hydrophobic Tail	HC	HC	hl	-0.0011
4 <sup>th</sup> CH <sub>2</sub> Hydrophobic Tail	CT	C8	a3	-0.0011
4 <sup>th</sup> CH <sub>2</sub> Hydrophobic Tail	HC	HC	hl	-0.0077
5 <sup>th</sup> CH <sub>2</sub> Hydrophobic Tail	CT	C8	a3	0.0220
5 <sup>th</sup> CH <sub>2</sub> Hydrophobic Tail	HC	HC	hl	-0.0114
6 <sup>th</sup> CH <sub>2</sub> Hydrophobic Tail	CT	C8	a3	0.0185
6 <sup>th</sup> CH <sub>2</sub> Hydrophobic Tail	HC	HC	hl	-0.0069
7 <sup>th</sup> CH <sub>2</sub> Hydrophobic Tail	CT	C8	a3	0.0222
7 <sup>th</sup> CH <sub>2</sub> Hydrophobic Tail	HC	HC	hl	-0.0081
8 <sup>th</sup> CH <sub>2</sub> Hydrophobic Tail	CT	C8	a3	0.0138
8 <sup>th</sup> CH <sub>2</sub> Hydrophobic Tail	HC	HC	hl	-0.0099
9 <sup>th</sup> CH <sub>2</sub> Hydrophobic Tail	CT	C8	a3	0.0095
9 <sup>th</sup> CH <sub>2</sub> Hydrophobic Tail	HC	HC	hl	-0.0079
10 <sup>th</sup> CH <sub>2</sub> Hydrophobic Tail	CT	C8	a3	0.0079
10 <sup>th</sup> CH <sub>2</sub> Hydrophobic Tail	HC	HC	hl	0.0001
11 <sup>th</sup> CH <sub>2</sub> Hydrophobic Tail	CT	C8	a3	0.0393
11 <sup>th</sup> CH <sub>2</sub> Hydrophobic Tail	HC	HC	hl	-0.0038
12 <sup>th</sup> CH <sub>3</sub> Hydrophobic Tail	CT	C8	a3	-0.0833
12 <sup>th</sup> CH <sub>3</sub> Hydrophobic Tail	HC	HC	hl	0.0166

Table S2: List of the biochemical used in experiment

I.	GST-DI	his6-DI
Lysis buffer	30mM Tris pH7.4, 500mM NaCl, 20% glycerol, 43mM LDAO, 1mM DTT	30mM Tris pH7.4, 500mM NaCl, 20% glycerol, 43mM LDAO, 20mM imidazole, 1mM DTT
Wash buffer	30mM Tris pH7.4, 300mM NaCl, 10% glycerol, 2mM LDAO, 1mM DTT	30mM Tris pH7.4, 1M NaCl, 10% glycerol, 2mM LDAO, 1mM DTT and 50mM imidazole
Elution buffer	30mM Tris pH7.4, 300mM NaCl, 10% glycerol, 2mM LDAO, 25 mM glutathion, 1mM DTT	30mM Tris pH7.4, 300mM NaCl, 10% glycerol, 2mM LDAO, 500mM imidazole, 1mM DTT

## II. List of Figures

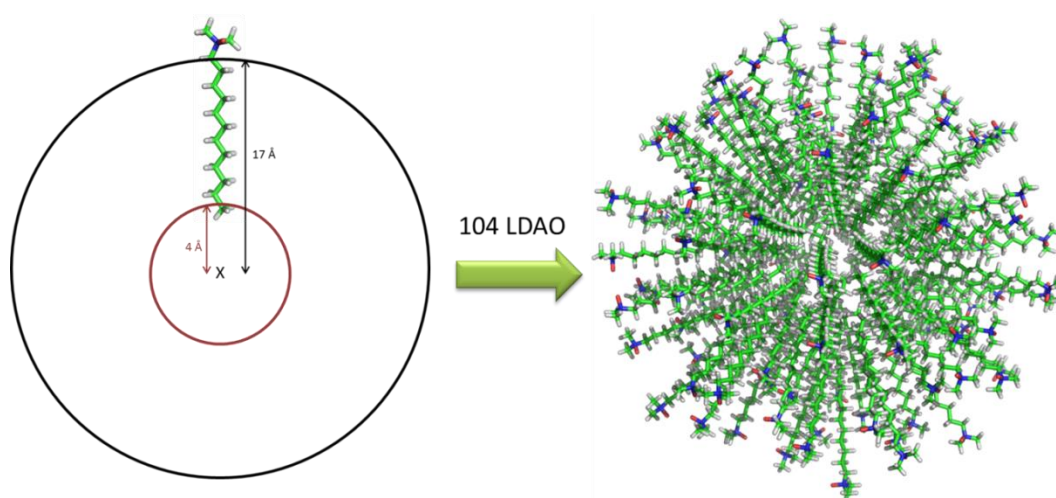


Figure S1: Geometric model used to construct the preformed LDAO with Packmol [38] see main text for details.



	1	11	21	31	41	
I-TASSER	DYSNFMASLN	LRNRYAGEVY	GMIRFSGTTG	QMSDLNKMMV	QDLHSALDRS	
RaptorX	DYSNFMASLN	LRNRYAGEVY	GMIRFSGTTG	QMSDLNKMMV	QDLHSALDRS	
Phyre2	DYSNFMASLN	LRNRYAGEVY	GMIRFSGTTG	QMSDLNKMMV	QDLHSALDRS	
GenThreader	DYSNFMASLN	LRNRYAGEVY	GMIRFSGTTG	QMSDLNKMMV	QDLHSALDRS	
PredictProt	DYSNFMASLN	LRNRYAGEVY	GMIRFSGTTG	QMSDLNKMMV	QDLHSALDRS	
QUARK	DYSNFMASLN	LRNRYAGEVY	GMIRFSGTTG	QMSDLNKMMV	QDLHSALDRS	
	51	61	71	81	91	
I-TASSER	HPQHYTQAMF	KLTAMLISSK	DCDPQLLHHL	CWGPLRMFNE	HGMETALACW	
RaptorX	HPQHYTQAMF	KLTAMLISSK	DCDPQLLHHL	CWGPLRMFNE	HGMETALACW	
Phyre2	HPQHYTQAMF	KLTAMLISSK	DCDPQLLHHL	CWGPLRMFNE	HGMETALACW	
GenThreader	HPQHYTQAMF	KLTAMLISSK	DCDPQLLHHL	CWGPLRMFNE	HGMETALACW	
PredictProt	HPQHYTQAMF	KLTAMLISSK	DCDPQLLHHL	CWGPLRMFNE	HGMETALACW	
QUARK	HPQHYTQAMF	KLTAMLISSK	DCDPQLLHHL	CWGPLRMFNE	HGMETALACW	
	101	111	121	131		
I-TASSER	EWLLAGKDGV	EVPFMREMAG	AWHMTVEQKF	G		
RaptorX	EWLLAGKDGV	EVPFMREMAG	AWHMTVEQKF	G		
Phyre2	EWLLAGKDGV	EVPFMREMAG	AWHMTVEQKF	G		
GenThreader	EWLLAGKDGV	EVPFMREMAG	AWHMTVEQKF	G		
PredictProt	EWLLAGKDGV	EVPFMREMAG	AWHMTVEQKF	G		
QUARK	EWLLAGKDGV	EVPFMREMAG	AWHMTVEQKF	G		

— Helix  
— Coil

Figure S2: Sequence alignment of DI protein with the secondary structure prediction of different methods. In red, the helix prediction and in green, the random coil prediction.

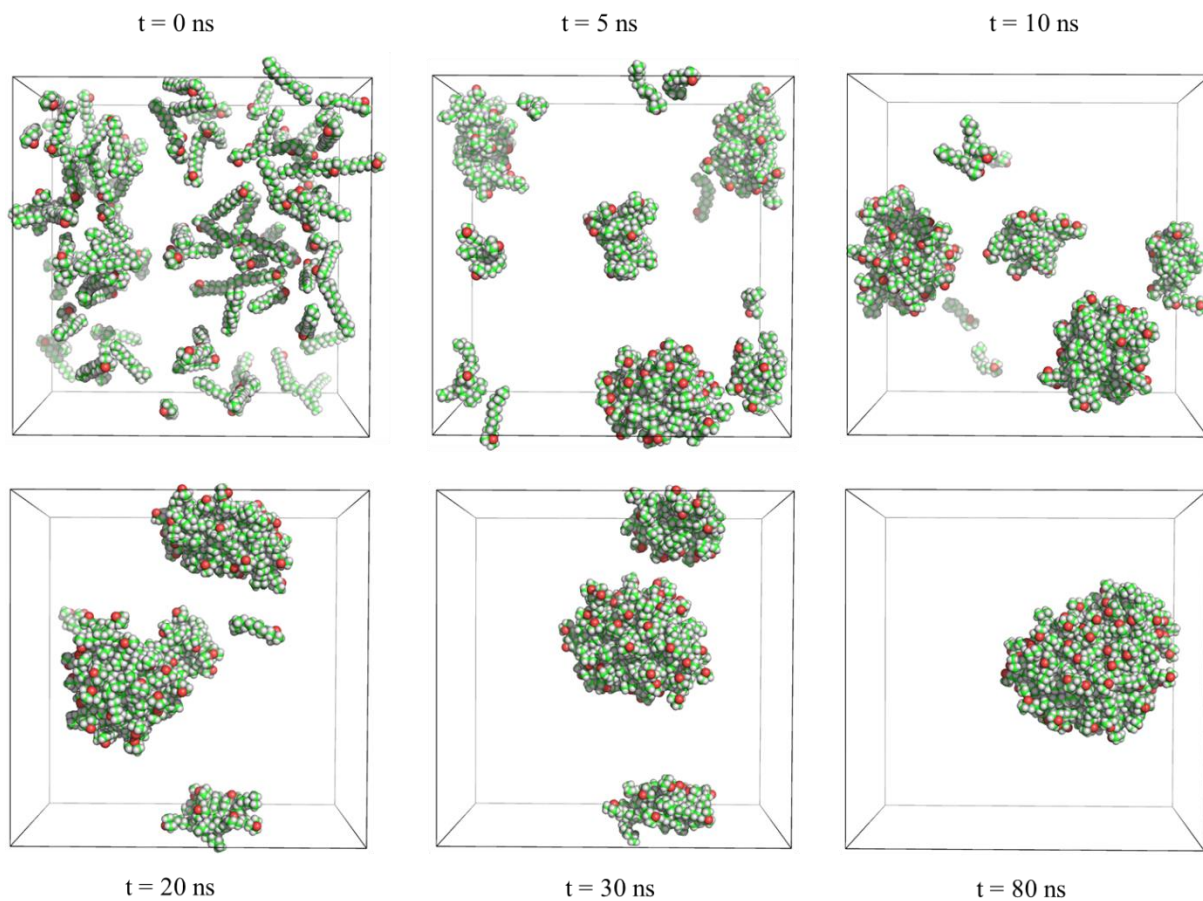


Figure S3: LDAO self-aggregation into a single micelle in GAFFLipid simulation [77].

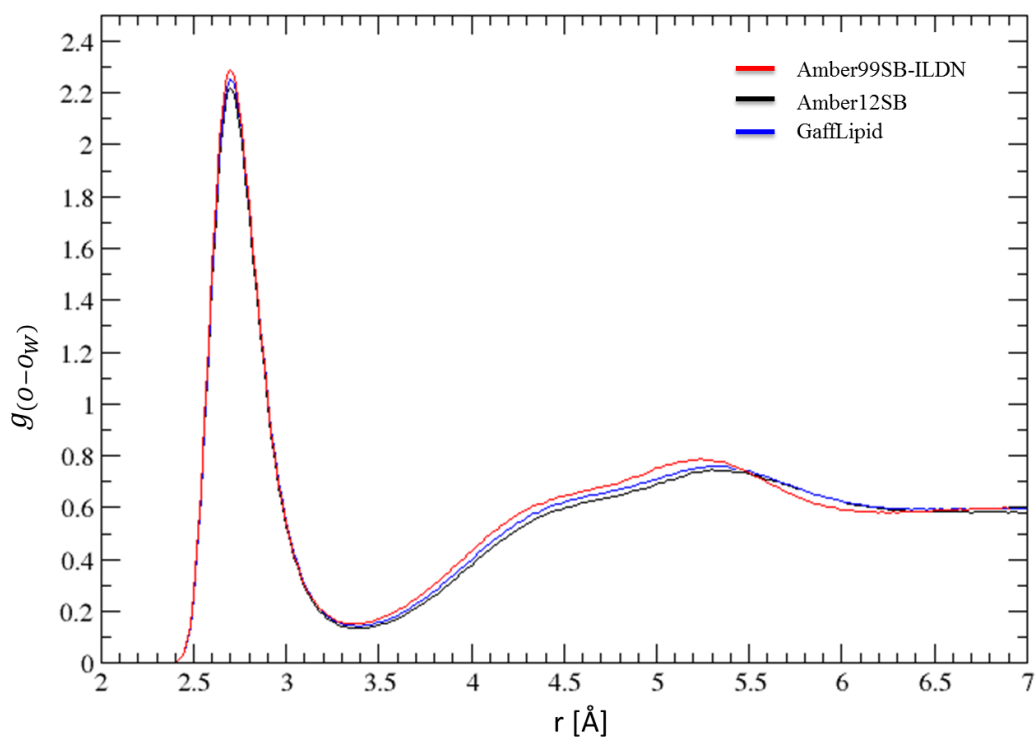


Figure S4: Average radial density function  $g_{(O-O_w)}$  between the N-oxide oxygen atoms (O) and water oxygen ( $O_w$ ) for the Amber99SB-ILDN (red line), Amber12SB (black line) and GaffLipid (blue line). The values for the SA and PF initial conditions are superimposed.

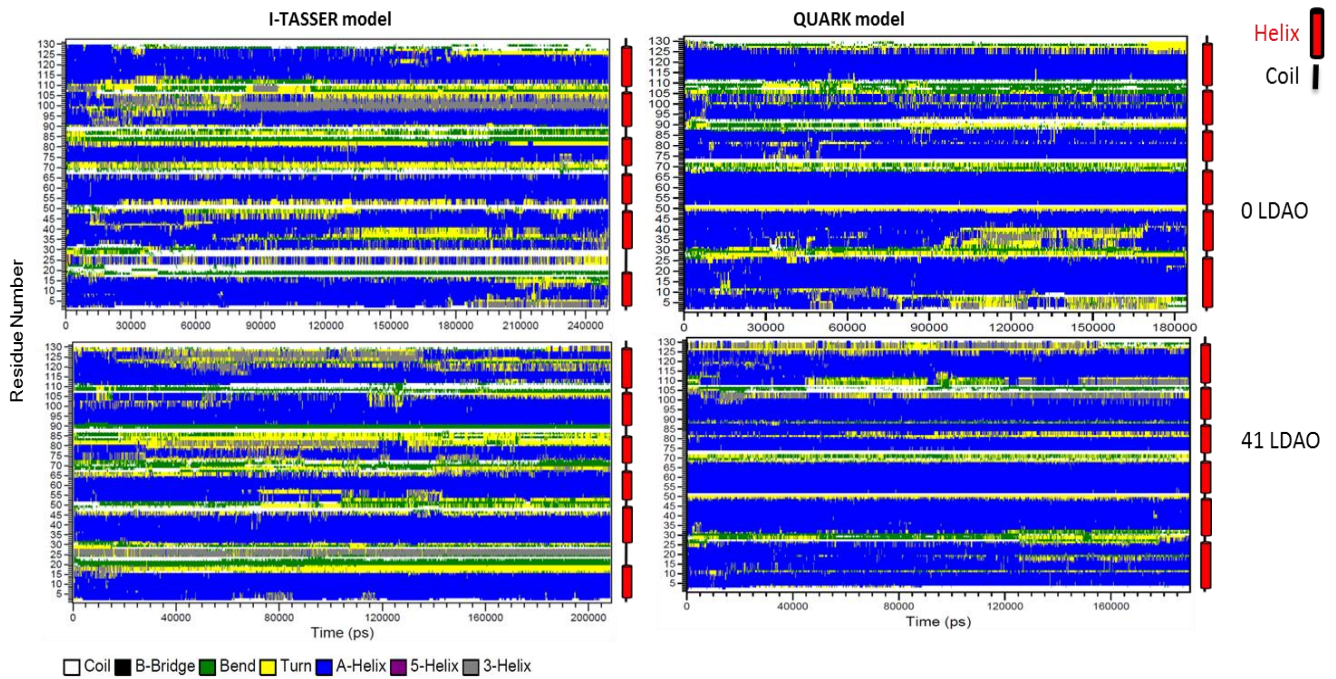


Figure S5: Secondary structure profile of DI protein as function of simulation time for simulation without and with LDAO.

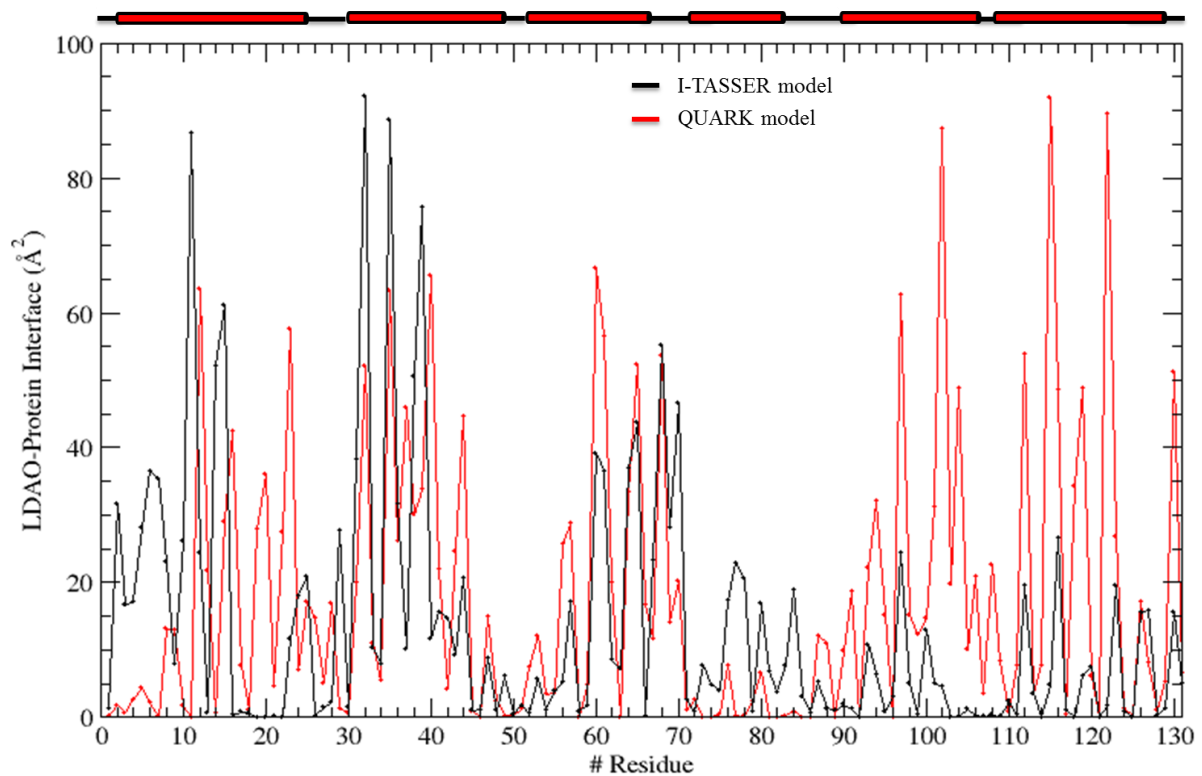


Figure S6: The LDAO-protein interfaces for all residues, in black for the I-TASSER model and in red for the QUARK model. We used *trjVoronoi* program to compute these interfaces in the last 100 ns.

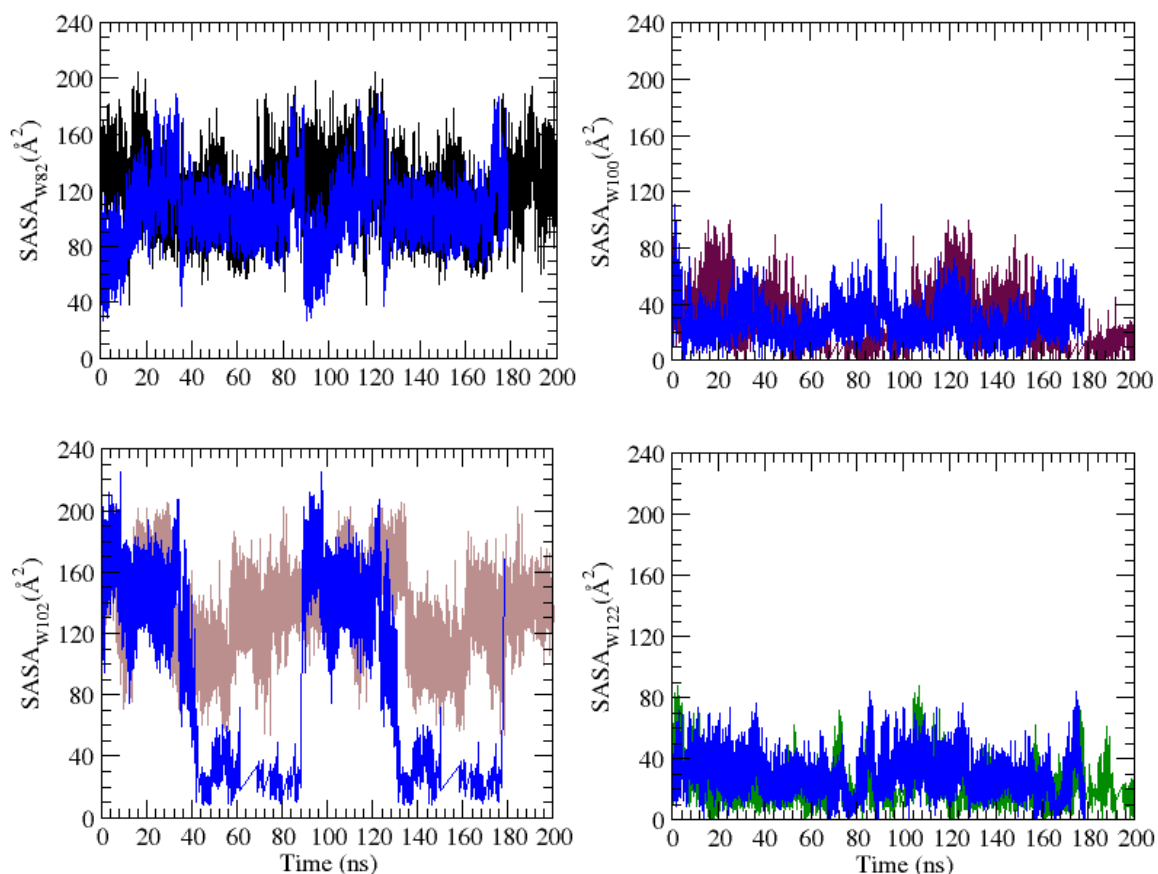


Figure S7: The SASA of tryptophan with the I-TASSER model. In black, in dark brown, in light brown and in green, the Trp82, Trp100, Trp102 and Trp122, respectively. In blue, the corresponding SASA of tryptophan without detergent. The Trp82, Trp100, Trp102 and Trp122 are buried 34 %, 57 %, 45 % and 83 % of simulation time, respectively.

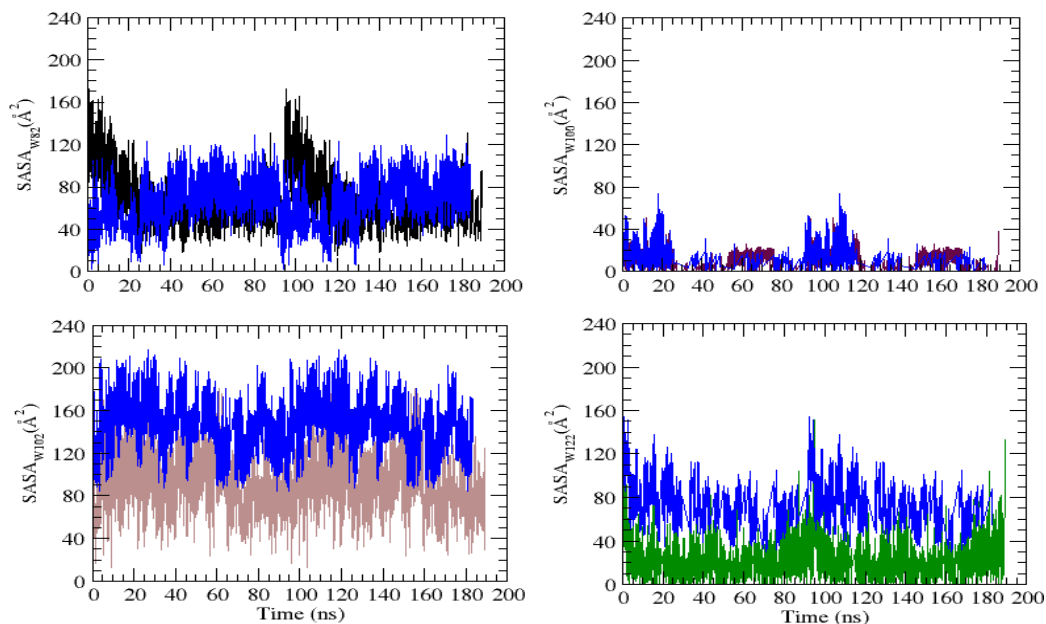


Figure S8: SASA of tryptophans with the QUARK model. In black, in dark brown, in light brown and in green, the Trp82, Trp100, Trp102 and Trp122, respectively. In blue, the corresponding SASA of tryptophan without detergent. The Trp82, Trp100, Trp102 and Trp122 are buried 48 %, 66 %, 97 % and 99 % of simulation time, respectively.

## 1) Caractérisations supplémentaires de DI et résultats préliminaires d'interaction avec d1-NS5A

### **i) Matériel et méthode**

#### ■ Caractérisation :

##### **Dichroïsme circulaire**

*Cette expérience a été réalisée avec l'aide Christophe Velours, ingénieur de recherche sur la plateforme IMAGIF de l'I2BC.*

La construction His-DI a été tout d'abord dialysée sur la nuit dans un tampon 50mM potassium phosphate pH 7.4, 300 mM NaF, 10% glycérol, 2mM LDAO dans des bouchons de dialyse de cut off 3.5kDa afin de diluer au maximum le NaCl non compatible avec les expériences de dichroïsme circulaire. Au total, 1ml d'échantillon a été mis en dialyse dans 1L de tampon. Les mesures de dichroïsme circulaire ont été effectuées dans un spectropolarimètre Jasco-810 équipé d'un système Peltier Jasco PTC-423S/15 pour le maintien de la cellule et de la température. 200µL d'échantillon à 0,07mg/mL ont été placés dans une cuve de trajet optique de 1mm. 10 acquisitions du tampon et de la protéine ont été enregistrées entre 185 et 260nm. Le bruit de fond étant important, seul le spectre entre 200 et 240nm a pu être exploité. En effet, entre 185 et 197 nm, la haute tension du photomultiplicateur est supérieure à 600V. Or, elle correspond à l'augmentation du voltage appliquée au photomultiplicateur qui va permettre d'augmenter le signal obtenu. Au-delà de 600V, le signal enregistré n'est plus considéré comme fiable. Les données ont été traitées par K2D2, seul logiciel autorisant à traiter les données entre 200 et 240 nm et non pas entre 190 et 240nm. En effet, il est des fois difficile d'obtenir des données de bonnes qualités avant 200nm, ce qui est notre cas ici puisque nous avons un fort bruit de fond. Une estimation du spectre basée sur la prédiction de structure secondaire a été réalisée au préalable par K2D2<sup>135</sup>.

##### **Pontage au glutaraldéhyde**

500µl de his-DI à 0,2 mg/mL a été au préalable dialysé dans une litre de 30mM Hepes pH7.4, 300mM NaCl, 10% glycérol, 2mM LDAO, 3mM DTT sur la nuit. En effet, le Tris peut altérer les pontages au glutaraldéhyde. Le lendemain, il a été ajouté à 100µl de protéine 5µL de glutaraldéhyde 2,3%. Le tout est incubé pendant 10 minutes à température ambiante puis la réaction a été arrêtée avec 10µL de Tris-HCl 1M pour 20µL de la solution. Les échantillons ont ensuite été déposés sur gel Tris-Tricine 15%.

##### **Chromatographie d'exclusion**

500µL de His-DI a été injecté sur S200 10/300 GL à un flux de 0,3mL.min<sup>-1</sup>. La colonne a été équilibrée au préalable dans le tampon de la protéine même : 50mM Tris pH 7.4, 300 mM NaCl, 10% glycérol, 2mM LDAO. Des fractions de 1 mL ont été collectées puis déposées sur gel Tris-Tricine 15%.



- Interaction avec le domaine d1 de NS5A

### **Expression et purification du domaine d1 de NS5A**

*Ce protocole a été mis au point par Isabelle Auzat, chercheur CNRS au sein du Laboratoire de Virologie Moléculaire et Structurale de Gif-sur-Yvette, maintenant département de virologie de l'I2BC.*

Le domaine 1 de NS5A de la souche H77 du VHC a été exprimé avec une étiquette hexahistidines en N-terminal. Cette construction comporte la séquence des résidus 28 à 222, et ne comporte donc pas l'hélice amphipatique N-terminale de la protéine virale. Cependant cette construction possède les résidus impliqués dans l'interaction avec PI4KA, soit les résidus 202-210 de NS5A<sup>28</sup>.

La protéine est exprimée en BL21(DE3) sur la nuit à 18°C dans un milieu minimum LB implémenté de 50µg/mL de kanamycine. L'induction est effectuée à DO<sub>600nm</sub>=0,7 avec 1mM d'IPTG. Le culot de 300 ml de culture est repris dans 10 mL de tampon de lyse 50 mM Tris pH 7,5, 150 mM NaCl, 0.1% triton X100 et une tablette de cocktail d'anti-protéases. La lyse a été faite par sonication de 3 fois 1 minute d'amplitude 30, impulsion 3. Deux centrifugations successives sont effectuées afin de garder uniquement la protéine soluble, puis le surnageant est chargé sur une colonne d'affinité pré-packée (HisTrap FF 1mL) au préalable équilibrée dans 50mM Tris pH8, 500 Mm NaCl, 30mM imidazole. La purification est réalisée sur Akta Purifier. Le lavage est effectué avec un gradient de 30 à 200 mM imidazole puis la protéine est éluée avec un second gradient de 200 à 500 mM imidazole. L'imidazole est ensuite retiré par dialyse.

### **Pulldown entre PI4KA-DI et NS5A-d1**

Plusieurs conditions ont été testées mais n'ayant obtenu que des résultats négatifs, seule une expérience de pull down est exposée dans le manuscrit. Cependant, les différents tests réalisés sont répertoriés en annexe 10 (GST/His pulldown, tampons, détergent, temps d'incubation).

DI (témoin1), NS5A (témoin2) et le mélange DI/NS5A(test interaction) ont été chacun incubé sur la nuit avec 20µL de résine glutathion Sepharose 4B sous agitation à 4°C. Chacune de ces solutions a été complétée avec le tampon du partenaire afin d'atteindre 200µL de volume final (voir tableau ci-dessous). Le lendemain, le surnageant a été retiré puis la résine a été lavée avec 200µL de tampon de d1-NS5A. La résine et la fraction non retenue ont ensuite été déposées sur gel Tris-Tricine 15%.



<b><u>Témoin 1</u></b>	<b><u>Témoin 2</u></b>	<b><u>Test interaction</u></b>
100µL GST-DI (0,06 mg/mL)	100µL de d1-NS5A (0,35mg/mL)	100µL GST-D1
+ <b>100µL tampon d1-NS5A</b> 25mM Hepes pH 7,5 100 mM NaCl 20% glycérol	+ <b>100µL tampon DI</b> 30mM Tris pH 7,4 300mM NaCl 10% glycérol 2mM LDAO	100µL d1-NS5A

### **Pontage au glutaraldéhyde**

Comme pour les expériences d'interaction par pulldown, le protocole énoncé est l'un des protocoles suivis parmi des nombreux tests. La base du protocole reste le même, seules les conditions annexes varient. Les témoins correspondent aux témoins énoncés dans le protocole de pull-down.

50µL de His-DI à 0,12mg.mL<sup>-1</sup> au préalable dialysé dans un tampon contenant de l'Hepes sont incubés avec 50µL de NS5A à 0,35mg.mL<sup>-1</sup> pendant une heure à température ambiante. 5µL de glutaraldéhyde à 2,3% sont ajoutés à la solution protéique puis le tout est incubé 5 minutes à 37°C. La réaction est arrêtée avec 10µL de Tris pH 8,3 1M puis 15µL de chacun des échantillons sont déposés sur gel Tris-Tricine 15%.

## ii) Résultats

Les modèles générés par I-Tasser et Quark ont pu être en partie validés par des expériences dichroïsme circulaire. Cependant, les données ne sont pas d'assez bonne qualité pour être publiées. En effet, le spectre n'est exploitable qu'à partir de 197nm, et ne permet donc pas de vérifier tous les éléments de structures secondaires possibles (pic positif à 190nm des hélices alpha, hélices poly-proline). Néanmoins, nous avons pu constater dans un premier temps que le spectre de déconvolution correspond au spectre prédit par K2D2, validant ainsi les prédictions de structures secondaires sur le fragment DI.

La protéine est majoritairement composée d'hélice  $\alpha$  (48%) avec deux minima à 208 et 222nm qui sont caractéristiques des protéines hélicales. Nous sommes donc éloignés des données obtenus par modélisation (73% d'hélice pour I-TASSER ; 83% pour Quark). Cela peut s'expliquer par le bruit de fond important lors de nos analyses en dichroïsme circulaire, nous menant ainsi à exploiter seulement une partie du spectre émis par des les hélices  $\alpha$  (200-240nm contre 190-240 nm typiquement) ainsi qu'à l'utilisation d'un seul des nombreux logiciels de déconvolution disponibles.

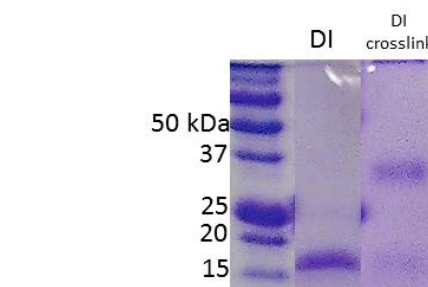
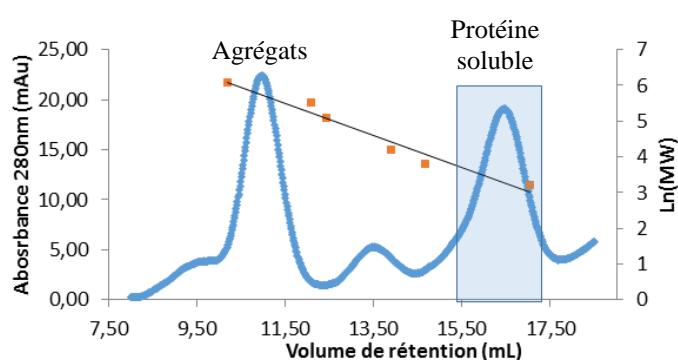
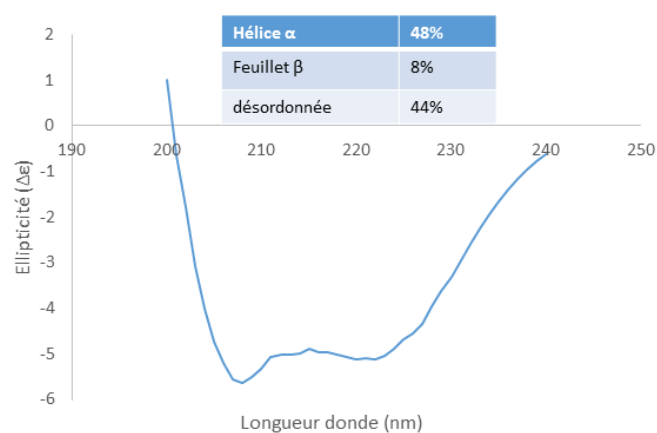
De plus, le minimum à 208 nm est plus intense que celui à 222 nm, à l'inverse des protéines uniquement composé d'hélice  $\alpha$ . Cela est caractéristique des protéines  $\alpha+\beta$ . Or, seulement 8% de brin  $\beta$  est prédit, soit 5 résidus sur les 153 composant DI, ce qui est trop peu pour former un feuillet  $\beta$ . De plus, le rapport  $\Theta$  ( $220/208$ ) = 0,9, ce qui est très proche du ratio rapporté pour les protéines composé majoritairement d'hélices  $\alpha$ . Mais il se pourrait qu'un autre élément apporte une contribution négative au pic négatif observé à 208 nm :

- les protéines majoritairement composées d'hélices 3 :10 ont un ratio  $\Theta(220/208)=0,5$ . Or, ces hélices sont assez courtes, ce qui pourrait être en accord avec la prédiction d'une petite hélice des résidus 85 à 89. De plus, lors des simulations de dynamique moléculaire, certains N- et C-terminaux d'hélice  $\alpha$  forment transitoirement des hélices 3 :10.
- Il semblerait que DI interagisse avec une molécule absorbant à 260nm. Ne connaissant pas sa nature ni son environnement exacte dans la protéine, nous ne pouvons rien affirmer quant à sa contribution dans le spectre. Cependant, cet élément pourrait être à l'origine de l'inversion des deux minima

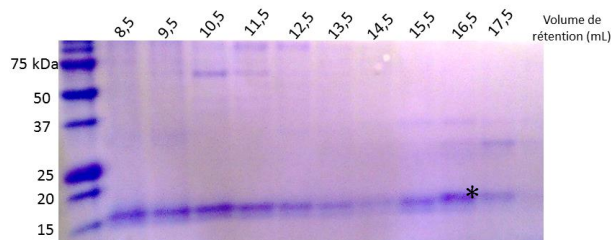
Un autre élément à valider afin d'établir de bonnes bases concernant les simulations de dynamique moléculaire était l'état d'oligomérisation de la protéine. Pour ce faire, deux études ont été menées : une étude sur gel filtration et une expérience de pontage. Grâce au SEC-MALLS, nous savions déjà que deux molécules de DI sont associées à 18.9 kDa de LDAO. Théoriquement, si DI est un monomère, nous devrions obtenir une protéine de 17,6 kDa en interaction avec au minimum 9,45kDa de LDAO, soit un complexe de 27kDa minimum lors des expériences de tamis moléculaire. Or, le volume d'élution de 6his-DI (Fig. 40) correspondrait à une protéine globulaire de 26 kDa, soit moins qu'un dimère de 6his-DI (35 kDa). Ainsi, le fragment DI serait un monomère en solution. Les simulations de dynamique moléculaires ont pu être donc menées sur les modèles monomériques générés par threading.

Cependant, nos résultats de pontage sont contradictoires puisque nous observons un déplacement net de la bande en SDS-PAGE (Fig. 40). Cependant, les résultats de pontage sont moins fiables que les résultats obtenus par gel filtration. En effet, le glutaraldéhyde forme des liaisons covalentes entre amines libres (lysine et N-terminal). Il s'avère que le N-terminal de DI est facilement accessibles, puisque on y trouve l'étiquette hexa-histidine et le site de coupure à la TEV, soit au minimum 21 résidus désordonnés.

Ainsi, même si la molécule de glutaraldéhyde est relativement courte, des faux positifs sont facilement obtenus. De plus, nous avons solubilisé DI en présence d'une grande présence de LDAO. Nous avons ainsi peut être des micelles mixtes LDAO-lipides. Or, les phospholipides peuvent réagir avec le glutaraldéhyde. Enfin, la bande obtenue est moins intense que le dépôt sans pontage. Une partie de la protéine a donc pu agréger lors de l'incubation à 37°C, favorisant ainsi la formation transitionnelle de dimère.



Protéine	MW (kDa)	Volume de rétention (mL)
1- ferritine	440	10,21
2- catalase	232	12,08
3- aldolase	158	12,43
4-conalbumine	75	13,89
5- albumine	67	14,67
6- chimotrypsinogène	25	17,05
His-DI	26,2	16,46



**Figure 40-** Caractérisation du domaine putatif DI par dichroïsme circulaire, pontage et chromatographie d'exclusion.

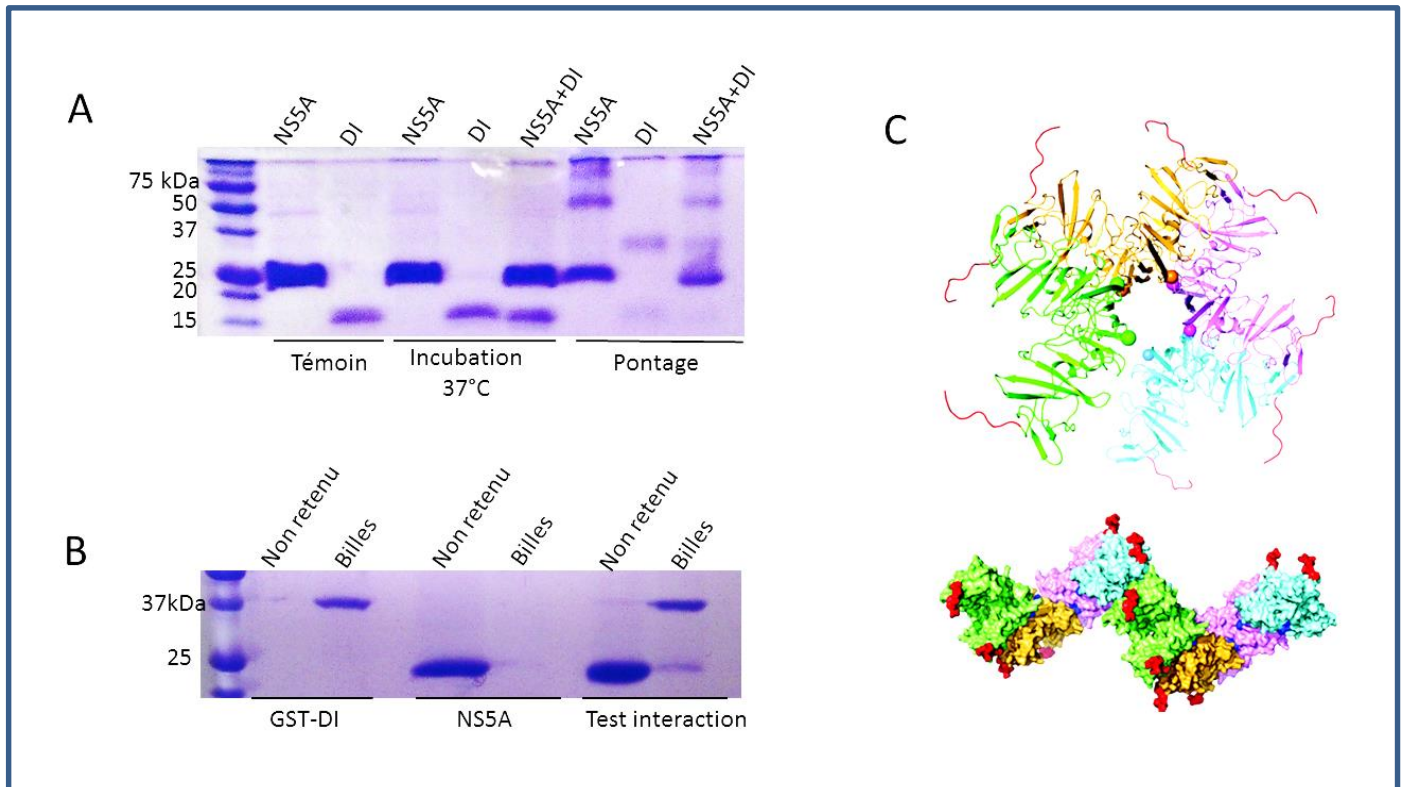
**En haut à gauche :** Spectre de dichroïsme circulaire de 6His-DI compris entre 197 et 240 nm.

L'estimation en structure secondaires calculée par K2D2 a été ajoutée dans le tableau placé au-dessus du spectre.

**En haut à droite :** Expérience de pontage avec 2,3% de glutaraldéhyde.

**En bas :** Expérience de tamis moléculaire sur Superdex 200 10/300 GL avec SDS-PAGE tris-Tricine 15% correspondant. Le chromatogramme a été obtenu pendant la purification de 6His-DI, ce qui explique pourquoi le pic correspondant au volume mort est aussi élevé ainsi que la présence d'un contaminant aux alentours de 13,5mL de volume de rétention. Chacun des marqueur orange correspond aux marqueur de poids moléculaire (répertoriés dans le tableau à droite) nécessaires afin d'évaluer le poids moléculaire théorique de 6His-DI (  $MW_{apparent}=26,2$  kDa ).

Le fragment DI est la région de PI4KA le plus susceptible d'interagir avec le domaine 1 de NS5A au vu des résultats obtenus par nos collaborateurs. Nous avons ainsi entrepris de vérifier si cette interaction était visible *in vitro*. Pour cela, des expériences de co-expression, co-purification, gel filtration, pull-down et pontage ont été réalisées. Entre nos mains, aucune interaction n'a été mise en évidence. Les résultats de deux de ces expériences sont exposés ici : expérience de pull down et de pontage au glutaraldéhyde.



**Figure 41-** Tests d'interaction entre le domaine DI de PI4KA et le domaine d1 de NS5A

**A. Expérience de pontage au glutaraldéhyde 2,3%.** Aucune interaction n'est visible (complexe 1 :1 = 38 kDa) cependant nous pouvons noter encore une fois la présence d'une bande correspondant à un dimère de 6his-DI ainsi que la présence de dimère et de tétramère du domaine d1 de NS5A.

**B. Expérience de GST-pull down sur la nuit.** Nous pouvons remarquer qu'une faible quantité de D1-NS5A semble interagir avec GST-DI dans ces conditions. Les expériences d'interaction sur une à trois heures ont quant à elles données des résultats négatifs.

**C (tiré de 136).** Plusieurs interfaces à l'origine de la dimérisation du domaine d1 de NS5A ont été résolues par cristallographie des rayons X. Un modèle d'oligomérisation prenant en compte deux d'entre elles a été proposé.

En ce qui concerne les résultats de pontage au glutaraldéhyde (figure 41-A), que ce soit avec LDAO en dessus ou en dessous de la CMC, aucune interaction n'a été mise en évidence. Cependant, plusieurs états oligomériques du domaine d1 de NS5A sont observés en solution : une majorité de monomère mais aussi la présence de dimère et un peu de tétramère. Dans ces conditions, le domaine 1 de NS5A est donc hétérogène. Des structures cristallographiques de différents dimères ont déjà été résolues, mais pas d'états oligomériques supérieurs. Des trimères et tétramères ont déjà été observés par pull down <sup>137</sup> mais dans notre cas précis, il est mis en évidence des états oligomérique vraisemblablement formés sur la base d'un dimère. Or, Love et al<sup>136</sup> ont modélisé des états oligomériques dont la brique fondamentale est un dimère (figure 41-C).

Lors de nos expériences de pull down, aucune conclusion fiable n'a pu être tirée. En effet, dans la majorité des cas, aucun complexe n'est visible sur les résines de glutathion ou de nickel. Cependant, nous avons systématiquement un décrochage d'une partie de la protéine de fusion normalement en interaction avec la résine, que l'incubation avec le partenaire protéique se fasse avant ou après l'incubation avec la résine. Ainsi, nous n'avons pu tirer de conclusion car :

- La construction GST-DI a tendance à interagir avec la résine de nickel. Ainsi le pull down en présence de résine de nickel n'était pas viable malgré la présence de d1-NS5A et de GST-DI dans la fraction non retenue
- Les tests en GST pull-down ont donné un seul résultat positif (figure 41-B). Cependant la fraction de d1-NS5A est faible.
- L'étiquette GST est plus grande que le domaine DI lui-même (17kDa contre 25kDa). De plus, la GST est une protéine qui dimérise alors que le domaine DI de PI4KA est un monomère. Le site d'interaction entre les deux partenaires peut être donc masqué par la GST

Ainsi, nous ne pouvons conclure quant à l'interaction entre le domaine putatif DI et le domaine d1 de NS5A. Cependant, au vu du caractère très hydrophobe du domaine DI de PI4KA, il se pourrait qu'il interagisse avec la membrane et soit donc à proximité du domaine d1 de NS5A au sein de la membrane du réticulum endoplasmique. Je pense à l'heure actuelle qu'il serait plus judicieux de repartir sur des tests d'interaction avec la PI4KA humaine entière. Il s'agirait ainsi de s'assurer des régions impliquées dans cette interaction *in vitro* avant de tester des constructions plus courtes de PI4KA dans l'ordre d'affiner la détermination du/des sites d'interaction avec la protéine virale.





# Partie III

---

Protéine périphérique à la membrane :  
solubilité de PI4KA, interaction à la  
membrane et modèle complet



# I. Matériel et méthode

---

## A. Expression en baculovirus de PI4KA entière et de 1129-2102

L'équipe de Volker Lohman nous a fourni le plasmide nécessaire à l'expression de PI4KA entière. Ce plasmide, un pFastBacDual, possède aussi un gène rapporteur GFP permettant de vérifier l'infection des cellules d'insecte par le baculovirus recombinant. La PI4KA est exprimée sous la forme de protéine de fusion contenant une étiquette hexa-histidines en N-terminal et une étiquette HA en C-terminal. A partir de ce plasmide, une seconde construction, 6-His-1129-2102, a été clonée en pFastBacB. Cette construction possède donc tout le module C-terminal de la PI4KA humaine, de DI au domaine catalytique. Cette construction avait été au préalable testée *in vitro* par nos collaborateurs et s'était avérée insoluble.

### 1) Constitution du stock de baculovirus par système Bac-to-Bac<sup>138</sup>

#### i) Constitution du bacmide en *E.Coli* DH10Bac

Une première étape consiste à la transposition du gène d'intérêt dans un bacmide contenant le génome du baculovirus. Ce gène est placé en aval du promoteur de la polyédrine (pour les pFastBacB et pFastBacDual) surexprimée dans les cellules d'insectes lors de l'infection par le virus ; ou en aval du promoteur de p10 (pFastBacDual), protéine impliquée dans la morphogénèse des corps d'inclusion formés par la polyédrine. La transposition est réalisable grâce à un plasmide « Helper » à l'origine de l'expression de la transposase nécessaire à la transposition du gène d'intérêt via le transposon bactérien Tn7. La sélection des bactéries recombinantes se fait par sélection blanc/bleu.

100µL de DH10BAC sont mis à incuber avec 1.4ng de pFastBacDual pendant 1h30 dans la glace. La transformation des bactéries par le plasmide se fait par choc thermique classique suivi d'une incubation à 37°C pendant 5 heures. Les bactéries sont ensuite étalées sur boîte LB additionné de 50µg.mL<sup>-1</sup> de kanamycine (bacmide), 7µg.mL<sup>-1</sup> de gentamicine (plasmide donneur), 100µg.mL<sup>-1</sup> de tétracycline (plasmide helper), 100µg.mL<sup>-1</sup> X-Gal et 40µg.mL<sup>-1</sup> d'IPTG (sélection blanc/bleu) puis incubée à 37°C pendant 2 jours.

Les colonies blanches sont ensuite repiquées en milieu LB liquide plus antibiotiques puis incubées sur la nuit sous agitation à 37°C afin d'amplifier le bacmide recombinant.

Le lendemain, les bactéries ont été lysées en utilisant les solutions A1, A2 et A3 du kit Nucleospin Plasmid de Macherey Nagel, kit de mini-prep. Le bacmide a ensuite été purifié par précipitation à l'isopropanol puis vérifié par PCR.

## ii) Constitution du stock de baculovirus

La constitution et l'amplification du stock de baculovirus ont été réalisées en cellules SF21 adhérentes, lignée cellulaire issue d'ovaires du papillon *Spodoptera frugiperda*.

Dans un premier temps, 1 µg de bacmide et 6 µL de cellfectine sont incubés pendant 45 minutes à température ambiante. La solution est complétée avec le volume nécessaire de TC100 afin d'atteindre un volume final de 200 µL. Le TC100 est un milieu conçu pour la culture de certains types de cellules d'insecte. La cellfectine est une solution de liposomes formés de différents lipides cationiques permettant l'entrée du bacmide dans la cellule eucaryote. Le processus n'est pas bien défini, mais la cellfectine provoquerait un phénomène d'endocytose du bacmide en médiant l'interaction ADN/membrane cellulaire et en condensant le bacmide.

800 µL de ce mélange est ajouté à des cellules confluentes adhérentes ( $9.10^5$  cellules d'insecte dans un puits de 33mm de diamètre) dans 2 mL de milieu TC100 sans antibiotique. Après incubation pendant 5 heures à 27°C, le surnageant est retiré et remplacé par 2 mL de TC100 avec 1/100<sup>ème</sup> de sérum fœtal bovin et de pénicilline/streptomycine. Après 4 jours, la présence de baculovirus peut être vérifiée dans le cas de PI4KA entière par la production de GFP. Le surnageant contenant le baculovirus est alors récupéré et centrifugé 15 minutes à 500g pour éliminer les cellules d'insectes. Ce surnageant constitue le stock P1 de baculovirus produit. Il est donc à conserver à -80°C précieusement et à amplifier.

*Le protocole qui suit est le protocole suivi pour toutes les infections. Seuls la quantité de virus utilisée et le temps d'infection changent.*

L'amplification se fait tout simplement par infection des cellules SF21 par incubation de cellules confluentes avec 500 µL de stok P1 (ou P2 suivant le stock à amplifier) dilué dans un peu de milieu TC100 (1,5mL pour une flasque T75) plus sérum et antibiotique. Après une heure d'incubation à 27°C, 13mL de TC100+sérum+antibiotiques sont ajoutées. L'infection peut être suivie par fluorescence dans le cas de PI4KA entière, ou bien par l'observation de la morphologie des cellules. Après 48 heures, le surnageant contenant le baculovirus amplifié est récupéré, centrifugé 15 minutes à 500g et filtré. Si nécessaire, une seconde amplification est à effectuer.

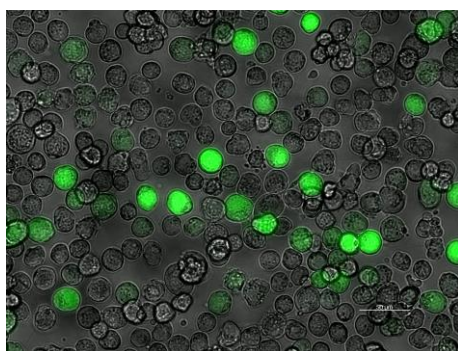
.

### iii) Test de Multiplicité d'Infection (MOI)

La MOI correspond au nombre de virus infectieux par cellules d'insecte lors de l'infection. Il est donc important de l'évaluer : une MOI trop forte entrainera une mort prématurée des cellules d'insectes alors qu'une MOI trop basse ne permettra pas une expression optimale de la protéine d'intérêt.

L'infection est réalisée dans des puits de 2mL contenant des SF21 confluentes. Le surnageant est retiré, puis sont ajoutés 50, 100 ou 500µL de baculovirus (MOI basse, moyenne et haute) avec ajout de TC100+sérum+antibiotique pour atteindre un volume final de 500µL. Trois infections sur trois puits différents sont effectuées. Les cellules sont collectées à 30, 48, et 72 heures. Pour cela, les cellules sont reprises dans 1 mL de PBS 1X puis centrifugées 5 minutes à 11600g.

Afin d'évaluer la quantité de protéine exprimée, soluble et insolubles, une lyse des cellules par congélation/décongélation est réalisée. Pour cela, les culots cellulaires sont repris dans 200µL de tampon 50mM  $\text{KH}_2\text{PO}_4$  pH 7,4, 300 mM NaCl, 1% Triton X-100 et ¼ de pastille de cocktail d'antiprotéases. Le culot et le surnageant sont séparés par centrifugation de 14000g pendant 10 minutes à 4°C puis sont déposés sur SDS PAGE 10%. La protéine est mise en évidence par western blot anti His.



***Figure 42-** Visualisation par microscopie à fluorescence de l'expression de la GFP au bout de 72heures post-infection sur cellules SF21*

Pour chacun des puits, les cellules ont été aussi analysées par microscopie à fluorescence afin de visualiser l'expression de la GFP. Grâce au nombre de cellules infectées et non infectées, nous avons pu calculer plus finement le titre du stock de baculovirus utilisé pour l'expression de la PI4KA entière. Elle a été estimée à  $1 \cdot 10^8$  pfu/mL pour une MOI moyenne. De plus, l'expression des PI4KA entière et de la construction 1129-2102 ont été vérifiées par Western Blot. Pour cela, il a été utilisé un anticorps primaire anti-His de souris (QIAGEN, ref 134660). La révélation a été réalisée grâce à un anticorps secondaire anti-souris (chèvre) couplée à la peroxydase.

#### iv) Choix de la souche de cellule d'insecte

Deux souches ont été testées : des cellules SF21, plutôt utilisées pour l'amplification des baculovirus ; et des H5, des cellules d'ovaire de *Trichoplusia ni* plutôt utilisées pour la surexpression des protéines. De plus, j'ai adapté en petit volume des SF21 adhérentes à la culture en suspension. Le même protocole est suivi pour l'infection de cellules SF21 et H5 adhérente. Seul le milieu est différent car les cellules H5 sont adaptées aux milieux sans sérum (milieu Ex-Cell).

Pour l'infection des SF21 en suspension, le protocole est le même que précédemment. Cependant, le baculovirus est directement dilué dans le surnageant contenant les cellules.

## B. Test de solubilité et interaction avec les membranes

### 1) Tests de solubilité de PI4KA entière et de 1129-2102

Afin d'évaluer la solubilité de la protéine, des tests de lyse ont été effectués avec différents tampons afin d'évaluer la solubilité de chacune des constructions en fonction de la concentration en NaCl, glycérol et détergent. Pour se faire, les culots cellulaires d'un puits de 33 mm de diamètre sont repris dans 200µL de tampon de lyse puis lysés par cinq cycles de congélation/décongélation. Les fractions solubles et insolubles sont ensuite séparées par centrifugation de 10 minutes à 18000g puis déposées sur SDS-PAGE 10%.

Premiers tests afin d'évaluer l'importance du glycérol :

50Mm KH<sub>2</sub>PO<sub>4</sub> pH 7,4+ 1mM PMSF + ¼ tablette anti protéase +

Tampon 1	Tampon 2	Tampon 3	Tampon 4	Tampon 5	Tampon 6
300 mM NaCl, 20% glycérol	500 mM NaCl	500 mM NaCl 20% glycérol	300 mM NaCl 5mM DTT	300 mM NaCl 5% glycérol	500 mM NaCl 5% glycérol

Seconds tests afin d'évaluer l'importance du détergent et du glycérol :

50Mm KH<sub>2</sub>PO<sub>4</sub> pH 7,4+ 1mM PMSF + ¼ tablette anti protéase +

Tampon 1	Tampon 2	Tampon 3	Tampon 4	Tampon 5	Tampon 6
300 mM NaCl, 1% Triton X-100	300 mM NaCl	300 mM NaCl 20% glycérol	300 mM NaCl 1% Triton X-100 20% glycérol	500 mM NaCl	500 mM NaCl 1% Triton X-100



## 2) Mise au point de la purification de PI4KA entière

Un premier protocole de purification de PI4KA entière avait été mis au point par nos collaborateurs (voir annexe 11). Cependant, la protéine était purifiée dans un tampon difficilement exploitable en biophysique : 50 mM  $\text{KH}_2\text{PO}_4$ , 150mM NaCl, 50% glycérol, 0,28  $\beta$ -mercaptoéthanol et 0,5% Triton X-100. Nous avons donc essayé d'évaluer l'impact du détergent sur la solubilité de PI4KA et d'abaisser le glycérol à 20% au cours de la purification.

Peu de conditions ont été testées par manque de temps, mais elles ont tout de même été répertoriées dans le tableau ci-dessous. Par ailleurs, un protocole a pu être mis au point avec une pureté suffisante pour réaliser les expériences dans de bonnes conditions.

<i>Lyse</i>	<i>[Triton X-100]</i>	<i>[NaCl]- glycérol</i>	<i>Histrap</i>	<i>Gel filtration</i>	<i>concentration</i>
1% triton 500 mM NaCl 20% glycérol Congélation / décongélation	0	500 Mm NaCl 10 % glycérol	OK	Superose6 10/300 GL Volume mort	perte
	0,006%	500 Mm NaCl 10 % glycérol	OK	/	Pas au delà de 0,2 mg/ml
	0,0155%	500 mM NaCl 10% glycérol	OK	Superose6 10/300GL OK	✓ (1 mg/mL)

Le culot bactérien de six boîtes de culture 10 cm de diamètre est repris dans 15 mL de tampon de lyse. Le tout est laissé à incuber 10 minutes dans la glace afin de fragiliser la membrane des cellules d'insectes. La lyse est ensuite réalisée par quatre cycles de congélation/ décongélation (27°C). Afin de s'assurer d'une lyse complète, le lysat est soniqué pendant 45 secondes à amplitude 45, impulsion 3. Les fractions solubles et insolubles sont séparées par centrifugation de 45 minutes à 40 000g à 4°C. Le surnageant est ensuite chargé sur colonne pré paquée de nickel de 1mL (Histrap 1mL HP) à un flux de 0,5 mL.min<sup>-1</sup>. Un premier lavage est effectué avec 0,5 % triton X-100 puis un second lavage avec 0,25 mM Triton X-100 afin d'éviter l'agrégation massive de protéine solubilisées grâce au détergent. L'élution est ensuite réalisée à 0,5 mL.min<sup>-1</sup> avec 300 mM d'imidazole. La protéine est collectée par fractions de 1mL.

**Tampon de lyse** : 50 mM  $\text{KH}_2\text{PO}_4$  pH 7,4 , 500 mM NaCl, 20% glycérol, 1 % Triton X-100, 1 mM DTT, 1 mM PMSF, 1 tablette de cocktail d'antiprotéases (sans EDTA)

**Tampon lavage n°1** : 50 mM  $\text{KH}_2\text{PO}_4$  pH 7,4 , 1M NaCl, 10% glycérol, 0,5% Triton X-100, 1 mM DTT, 25 mM imidazole.

**Tampon lavage n°2** : 50 mM  $\text{KH}_2\text{PO}_4$  pH 7,4 , 1M NaCl, 10% glycérol , 0,0155 % Triton X-100, 1mM DTT, 25 mM imidazole

**Tampon élution** : 50 mM  $\text{KH}_2\text{PO}_4$  pH 7,4 , 500 mM NaCl, 10% glycérol, 0,0155% Triton X-100, 1 mM DTT, 300mM imidazole

Cette purification n'est pas adaptée aux méthodes biophysiques car l'échantillon est encore trop contaminé, cependant la protéine est assez pure pour les expériences de flottaison de membrane et de protéolyse ménagée. Par manque de temps, le protocole n'a pu être terminé pour réaliser les analyses biophysiques escomptées.

Le Triton X-100 est un détergent qui possède un pic d'absorbance à 275 et 283nm. Nous ne pouvons donc pas réaliser une simple mesure d'absorbance à 280 nm afin d'évaluer la concentration de protéine. Pour se faire, nous nous sommes basés sur l'absorbance des liaisons peptidique à 205 nm et suivi la formule suivante :

$$[\text{protéines}] = A_{205\text{nm}}/31$$

Cette relation a été au préalable vérifiée par des mesures d'absorbance de différentes solutions de BSA de concentrations connues et avec différentes concentrations en Triton X-100.

La stabilité de PI4KA a aussi été testée en présence d'un second détergent, le N-octylglucoside. Pour cela, 500µL de PI4KA purifiée sans détergent a été concentrée dans 50µL avec un concentrateur de cut off 30 kDa à membrane PES. 0,1 ou 1% de NOG ont été ajoutés au préalable à la solution à concentrer. Le même test a été effectué avec 0,0155 % et 0,5% de Triton X-100.

### **3) Flottaison de membrane**

#### **Liposomes**

70µL de phosphatidylcholine à 10mg.mL<sup>-1</sup> sont mélangés à 30µL de phosphatidyl-éthanolamine à 10 mg.mL<sup>-1</sup>. Le solvant est évaporé à l'aide d'un système Speedvach pendant 30 minutes. Les lipides sont ensuite repris dans 500µL de tampon de la protéine dépourvu de détergent (50mM KH<sub>2</sub>PO<sub>4</sub> pH 7,5, 500 mM NaCl, 10% glycérol). La solution est ensuite passée au bain à sonication pendant 30 minutes puis incubée 15 minutes à température ambiante. Les liposomes sont enfin stockés à -20°C à l'abri de la lumière.

#### **Flottaison de membrane**

500µL de la préparation de liposome à 2mg.mL<sup>-1</sup> sont mis à incuber avec 100µL de protéines pendant 15 minutes à température ambiante. 500µL du mélange sont ensuite déposés au fond du gradient de sucrose composé de trois coussins : 2,5 mL de sucrose 80%(w/v) ; 5 ml à 50% (w/v) et enfin en surface 3mL à 10% (w/v). Le tout est mis à centrifuger à 35000g pendant 17 heures à 4°C. Ainsi, les liposomes, moins denses que le gradient de sucrose, vont remonter à la surface. Par conséquent, si PI4KA interagit avec les lipides, elle remontera avec les liposomes. A contrario, si cette interaction est inexistante, PI4KA restera au fond du gradient. Après centrifugation, des fractions de 1 mL sont collectées du bas vers le haut du gradient à l'aide d'une pompe péristaltique.

L'expérience n'a pu être réalisée qu'une seule fois. Il s'agit donc de résultats préliminaires. Cependant, ces résultats sont importants quant à la caractérisation de l'interaction de PI4KA avec les membranes.

#### 4) Tests de protéolyses ménagées

Afin de déterminer les domaines autonomes composants la PI4KA humaine entière ainsi que la stabilité de la protéine, trois expériences de protéolyse ménagée ont été menées :

- Avec la thermolysine, endoprotéase qui coupe en amont des résidus hydrophobes dans l'ordre suivant : Ile>Leu>Val>Phe. Cette coupure est rendu impossible par la présence de proline en amont ou en aval de ces derniers
- La chymotrypsine, qui coupe principalement en aval des résidus aromatiques Tyr, Phe, Trp
- La Subtilisine, endoprotéase non spécifique

Pour toutes ces expériences, plusieurs ratio protéase : PI4KA ont été testées : 0 ; 1 :10 ; 1 :30 ; 1 :100 ; 1 :300 ; 1 :1000 ; 1 :3000 ; 1 :10000 ; 1 :30000.

Dans les trois cas, protéines et protéases ont été mélangées dans le tampon de purification de la PI4KA, soit 50mM  $\text{KH}_2\text{PO}_4$ , 500 mM NaCl, 10% glycérol, 0,0155% Triton X-100, 1mM DTT dans un volume final de 25 $\mu\text{L}$ . La solution est incubée une heure à 37°C. La protéase est finalement inactivée avec 1 $\mu\text{L}$  d'EDTA 0,5M. Chacune des réactions est finalement déposée sur SDS-PAGE 15% .

## II. Résultats

---

Afin d'évaluer la solubilité des constructions PI4KA entière et de 1129-2102, des tests de lyse avec différents tampons ont été effectués. Le culot étant parfois difficile à solubiliser, seuls les surnageants peuvent être comparés entre eux.

Dans un premier temps, nous avons pu constater que la PI4KA entière n'a pas besoin de détergent pour être soluble. Cependant le seuil de solubilité de PI4KA dans ces conditions reste très faible puisqu'il est de 0,2mg/mL. Les meilleurs résultats ont été obtenus dans un tampon à forte force ionique (300 à 500 mM NaCl) avec 20% de glycérol. Le Triton X-100 n'a pas permis de solubiliser plus de protéine. Ainsi, dans le meilleur des cas, 50% de la PI4KA exprimée cellules d'insecte est soluble. Ce ratio est tout de même meilleur que celui obtenu pour les constructions exprimées en bactérie, pour lesquelles la majorité de la protéine est exprimée sous forme de corps d'inclusion. La solubilité de PI4KA entière exprimée en cellules eucaryotes est ainsi supérieure aux constructions exprimées bactérie.

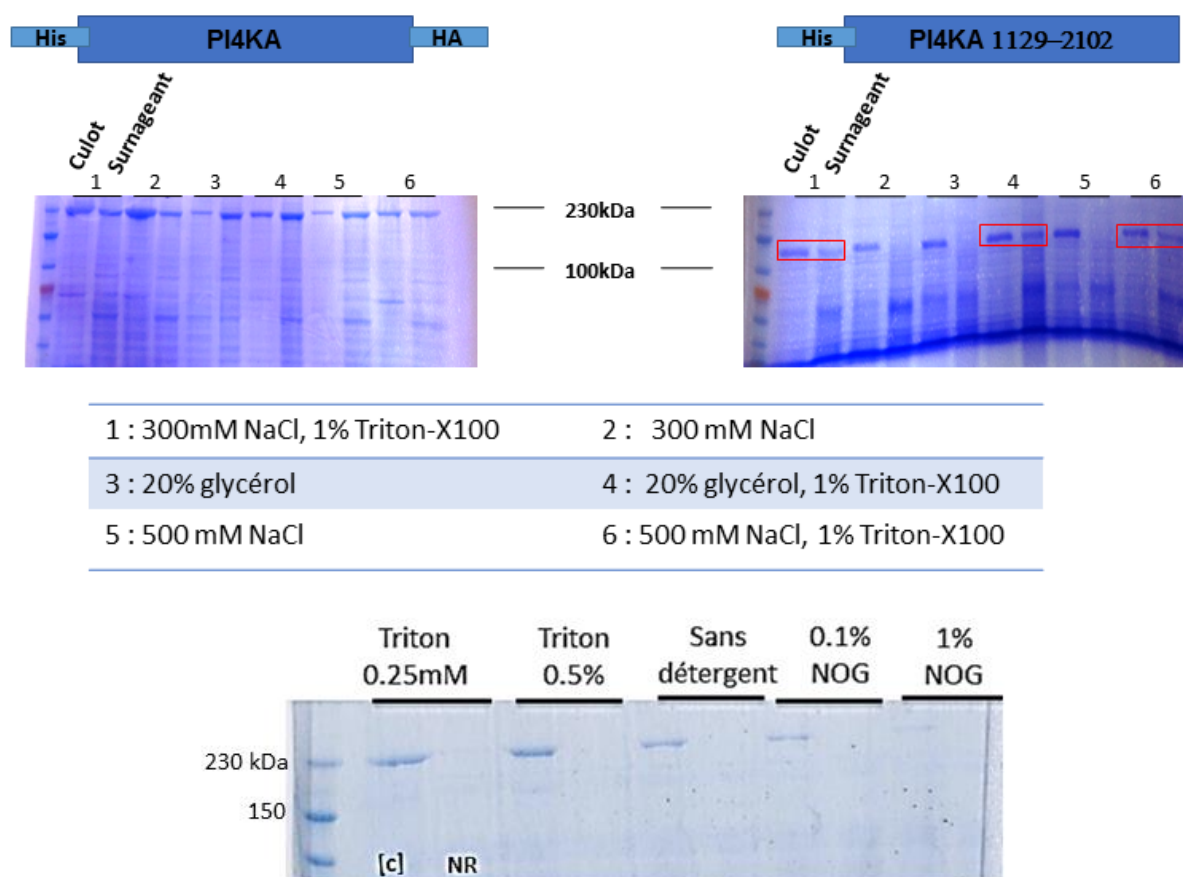
Globalement, la PI4KA entière a besoin des mêmes éléments chimique que le domaine N-terminal pour une stabilité optimum. Tout comme pour 26-1118, le DTT a aussi un effet sur la PI4KA entière. En effet, le tampon sans glycérol mais à 500mM NaCl et 5 mM DTT permet une solubilisation de la protéine aussi efficace qu'en présence de 20% glycérol. Au-delà des 45 cystéines présentes dans la séquence primaire de la PI4KA humaine susceptibles de former des ponts disulfures inter-macromoléculaires, la PI4KA a besoin de DTT pour être active<sup>59</sup>. Trois cystéines du domaine catalytiques ont été déjà étudiées :

- Les C1897 et C1901, des cystéines proches du cycle de l'adénine. Le mutant C1901S n'inactive pas la kinase. Dans la modélisation publiée dans l'article n°1, ces deux cystéines sont enfouies. Ces cystéines semblent plutôt impliquées dans le bon repliement de la protéine. Seule C1901 est strictement conservée entre PI4KA et STT4p.
- La C1854, conservée entre PI4KA et STT4P mise à part chez les insectes pour lesquels la cystéine est substituée par une valine. Cependant, les mutations C1854L/S entraînent une perte d'activité kinase de la protéine. Encore une fois, cette cystéine est enfouie dans le lobe N-terminal et interviendrait donc dans la bonne stabilité de la protéine. Cependant, cette cystéine est située dans un motif strictement conservé entre PI4KA/STT4p 1847- AIFKVGDDCRQD-1858 mais pour lequel aucune fonction n'a été déterminée.

Les kinases sont connues pour être des protéines aux réarrangements conformationnels importants pour leur activité, ce qui expliquerait pourquoi la mutation de cystéine enfouie en sérine inactiverait la protéine. Cependant, d'après les modèles par homologie réalisés, ces cystéines ne semblent pas accessibles aux agents réducteurs.

Au contraire, la cystéine 1893, résidu strictement conservé entre PI4KA et STT4p, et à proximité de l'insertion 1810-1840 susceptible d'intervenir dans la régulation de l'activité de PI4KA par phosphorylation. Au-delà du module C-terminal, le DTT semble aussi bénéfique sur le module N-terminal. Cependant, aucune cystéine n'est strictement conservée dans le module 1-1118 de la PI4KA.

Au-delà des tests de lyse, des tests de concentration sans détergent, avec Triton X-100 et NOG en dessous et en dessus de la CMC ont été réalisés afin d'évaluer la solubilité de la protéine concentrée. Il s'avère que le Triton X-100 permet une meilleure solubilité de PI4KA entière. A noter que le NOG est un détergent non ionique, tout comme le Triton X-100, mais forme des micelles plus petites, et est caractérisé par une CMC bien plus élevée (23 mM). Visiblement, ce détergent diminue fortement la solubilité de la PI4KA entière. Nous savons donc que certain(s) détergent(s) permettent une meilleure solubilité de la protéine certainement au dépend de la taille de la micelle mais peut être aussi de leur courbure.



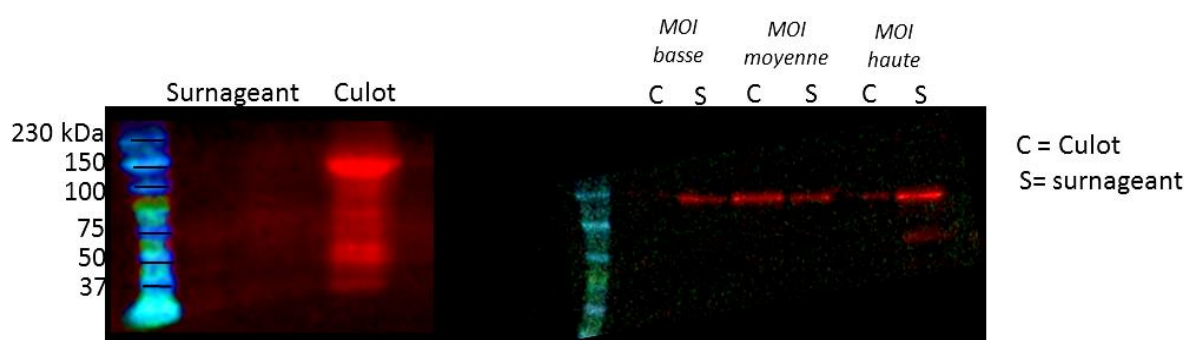
**Figure 43-** *Evaluation de la solubilité de la PI4KA humaine exprimée en cellules d'insecte SF21*

**En haut :** test de lyse de PI4KA entière et du module C-terminal dans différents tampons. La solubilité de PI4KA entière est bien plus élevée que celle de 1129-2102 qui est quasi nulle, même en présence de détergent.

**En bas à gauche :** Tests de concentration de PI4KA entière en présence de Triton X-100 ou de NOG. Le Triton X-100 favorise la solubilisation de la protéine alors que le NOG la défavorise fortement.

Lors de l'estimation de la MOI la plus favorable à l'expression de PI4KA et de 1129-2102, une fois de plus le caractère peu soluble de PI4KA a pu être mis en avant. Dans un premier temps, nous nous sommes rendus compte que peu importe la MOI, la construction 1129-2102 était encore une fois très peu soluble malgré l'expression plus forte de la protéine de fusion comparé à la PI4KA entière. De plus, lors de la vérification par Western Blot (**figure 44**), un smear est mis en évidence entre la protéine de fusion (110kDa) et le marqueur de poids moléculaire 50kDa où une double bande intense est observée. L'étiquette hexa-histidines se situant en N-terminal de la construction, la bande révélée à 50 kDa correspond au module putatif DI-DIII qui ne semble donc pas soluble non plus puisque détecté dans le culot après lyse. Encore une fois, cela est en accord avec nos études bio-informatiques ainsi que notre analyse du domaine DI. Cette construction semble ainsi être sujette à la protéolyse malgré la présence de PMSF. Par ailleurs, le caractère très hydrophobe de DI empêche de savoir si la construction n'est tout simplement pas soluble en Triton X-100 ou bien si la protéine est exprimée sous forme de corps d'inclusion. C'est pourquoi seule la purification de la protéine entière a été mise au point.

En ce qui concerne la PI4KA humaine entière, seule la MOI haute a permis une expression soluble convenable au bout de 72 heures. Le Western-Blot ne révèle pas de smear, cependant une bande apparaît au bout de 48 heures d'infection à MOI haute. Ce produit de dégradation soluble est visible entre les marqueurs de poids moléculaire de 150 et 100 kDa, ce qui correspond parfaitement au module N-terminal délimité. Par conséquent, les modules N- et C-terminaux seraient bien séparés par un linker accessible au solvant. Comme pour les tests de lyse, la solubilité moyenne de la protéine est d'environ 50%.



**Figure 44-** Tests de lyse des cellules d'insectes exprimant 1129-2102 et PI4KA entière- les protéines d'intérêt ont été révélées par Western Blot

**A gauche :** test de lyse de cellules infectées par le baculovirus permettant l'expression de 1129-2102 après 48h d'infection à MOI moyenne.

**A droite :** test de lyse de cellules infectées par le baculovirus permettant l'expression de PI4KA entière après 72h d'infection à différentes MOI.

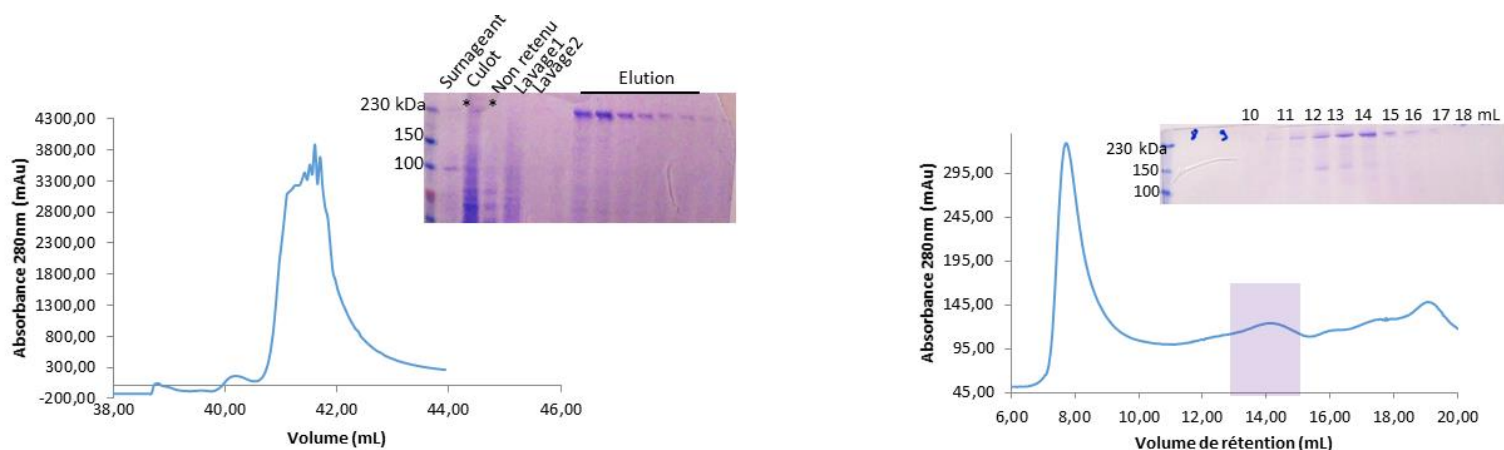


Lors de la mise au point de la purification de la PI4KA humaine, la présence de détergent au-dessus de la CMC s'est avérée essentielle pour la stabilité de la kinase. En effet, sans détergent la protéine agrège tandis que la protéine purifiée en Triton X-100 en dessous de la CMC possède un diamètre de 45,70 nm (intensité) par mesure en DLS. Le mélange est cependant homogène, sans agrégats massifs.

En sortie de gel filtration, PI4KA solubilisée dans le Triton X-100 à 0,25mM possède un poids moléculaire expérimental de 673,34 kDa et un rayon de Stokes  $R_s$  de 8nm. Or, le diamètre maximum  $D_{max}$  du module N-terminal calculé par SAXS est de 14,3 nm.

Dans le cas où la protéine n'interagit pas avec la matrice, le  $R_s$  est donc en accord avec un monomère. Cela implique une structure globale plutôt compacte, avec le module C-terminal stacké contre ce dernier, comme suggéré dans la partie précédente. Cependant, le nombre de molécule de détergent en interaction avec la protéine ne peut être déterminé de par la forme allongée de PI4KA. Pour cela, une expérience de SEC-MALLS est nécessaire.

Contrairement aux PI3Ks et à la PI4KIII $\beta$ , PI4KA a besoin de détergent pour être soluble au-delà de 0,2mg/mL et possède un caractère hydrophobe assez fort. Il semblerait donc que la PI4KA interagisse plus fortement avec les membranes que ce qui est supposé, ou que cette protéine ait besoin d'un partenaire protéique et/ou un co-facteur afin d'être plus soluble en solution. Il semblerait donc que PI4KA ait un mode de fonctionnement un peu différent que ses homologues proches.



**Figure 45-** Suivi de la purification de la PI4KA humaine

*A gauche :* Suivi de l'étape de chromatographie par affinité (nickel) à 280nm et SDS-PAGE 10% correspondant.

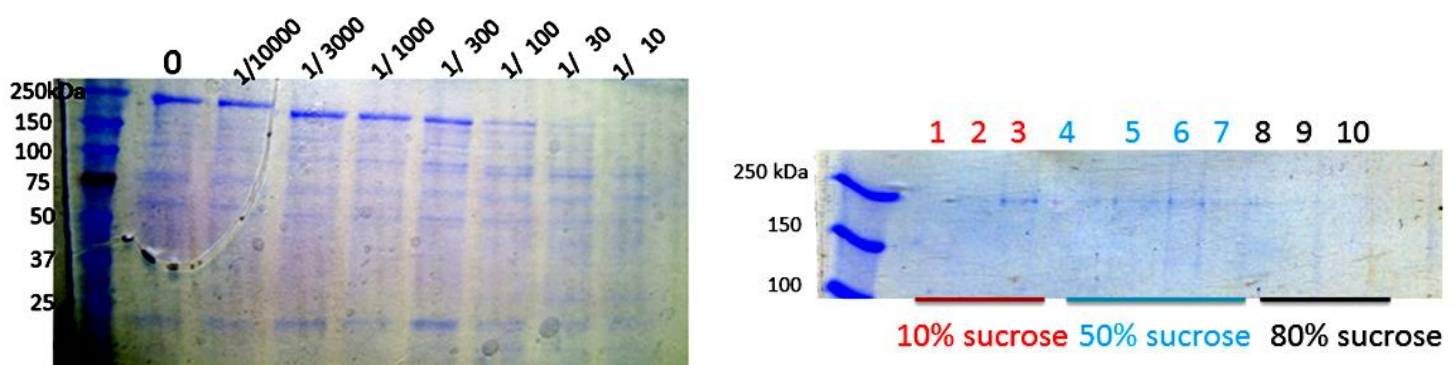
*A droite :* Profil d'élution de la PI4KA humaine injectée sur Superose 6 10/300 GL suivi à 280nm et SDS-PAGE 10% correspondant. Le Triton X-100 absorbant à 280nm, le profil d'élution ne rejoint pas la ligne de base, certainement en raison de son élution au niveau du volume totale de la colonne. PI4KA semble bien repliée en présence de détergent.



De manière surprenante, les expériences de protéolyses ménagées arborent les mêmes profils. En effet, un grand nombre de fines bandes apparaissent entre 230 et 75 kDa, sans qu'aucune ne se démarque vraiment. Le module N-terminal ne ressort pas clairement de ces expériences comme le laisserait penser les analyses des tests de MOI et la délimitation des modules de PI4KA. Ceci s'explique facilement pour la thermolysine et la chymotrypsine, endoprotéase spécifique, néanmoins ce n'est pas le cas de la subtilisine. Or, le domaine à répétitions HEAT ainsi que le domaine ARM possède des boucles accessibles au solvant de par leurs motifs répétés, ce qui pourrait expliquer ce profil atypique.

Un ratio protéase:protéine assez faible de 1 :1000 est requis pour un découpage efficace de la protéine par la subtilisine, attestant des nombreuses boucles accessibles au solvant. Des ratios plus faibles sont nécessaires pour les protéases spécifiques : 1 :300 pour la chymotrypsine et 1 :30 pour la thermolysine. Par conséquent, un certain nombre de résidus aromatiques sont en surfaces de la protéine, justifiant le caractère hydrophobe de PI4KA. Or, ces résidus sont connus pour favoriser les interactions avec les membranes. Les résidus hydrophobes moins volumineux quant à eux sont plus difficiles d'accès, ils sont donc plus enfouis que les résidus aromatiques.

De par son caractère fortement hydrophobe, nous nous demandions si PI4KA interagissait directement avec la membrane ou bien par l'intermédiaire de certains de ses partenaires protéiques. D'après nos résultats de flottaison de membrane, PI4KA interagit directement avec les liposomes, même s'ils sont dépourvus de PtdIns et constitué uniquement de phosphatidyléthanolamine et de phosphatidylcholine.



**Figure 46-** *Expérience de protéolyse ménagée par chymotrypsine (à gauche) et expérience de flottaison de membrane (à droite)*

Il s'agit de résultats préliminaires, ce qui explique la trainée de protéine visible tout au long du gradient de sucrose. Une centrifugation à plus grande vitesse ou plus longue sur la durée serait nécessaire pour un résultat plus net. Cependant, il est clair que PI4KA interagit, dans ces conditions, directement avec les liposomes. La protéine interagit donc avec les membranes sans même reconnaissance spécifique de PtdIns.

### III. Discussion : Similitude avec les PIKKs, Structure complète de PI4KA et fiabilité de la modélisation

---

#### **PI4KA .... Une PI4KKA ?**

Le domaine catalytique ainsi qu'une partie du domaine ARM des PI3Ks sont homologues aux PIKKs. Les PIKKs forment une grande famille de sérine/thréonine kinase de protéines interagissant avec les membranes. Seul Baretić et al en Décembre 2014<sup>79</sup> mentionnent l'homologie du domaine catalytique de PI4KA avec les PIKKs. Nos données et nos analyses suggèrent que cette homologie est bien plus étendue que ce qui était suggéré. Quand on parcourt les structures du C-terminus au N-terminus :

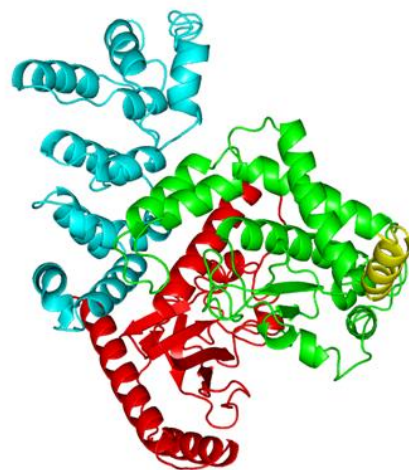
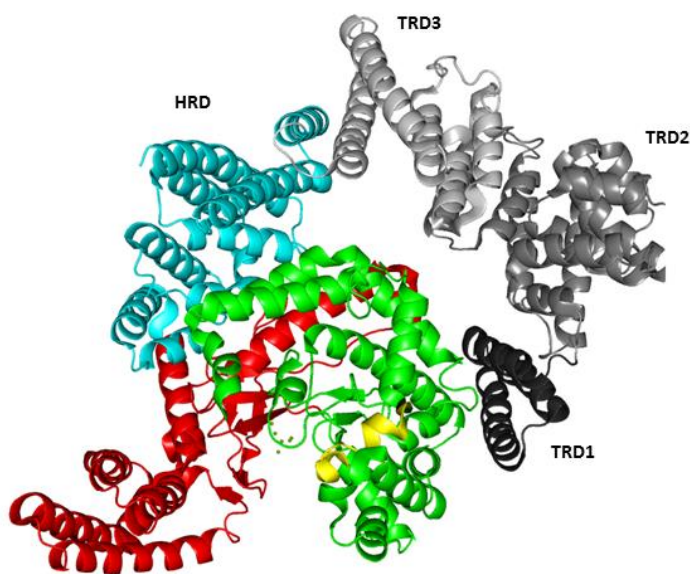
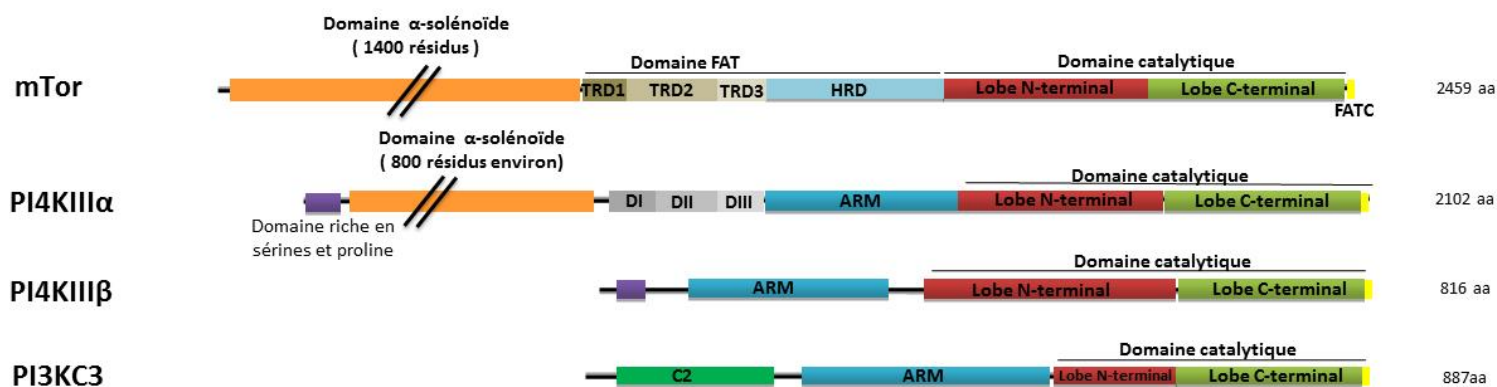
#### **- Domaine FATC des PIKKs VS hélice C-terminale de PI4KA**

L'hélice C-terminale des PIKKs, nommée FATC, est hydrophobe lorsqu'elle est isolée du reste de la protéine et lie la membrane *in vitro*. De plus, cette hélice est très conservée parmi les PIKKs et interagit avec la boucle d'activation, tout comme l'hélice C-terminale de la PI4KA/STT4p. Cependant, chez les PIKKs l'hélice FATC est masquée par une insertion très variable qui permet l'interaction avec des partenaires protéiques<sup>139</sup>.

#### **- Le domaine FAT vs DIII-ARM**

Le domaine FAT est peu conservé entre les PIKKs mais est nécessaire à leur activité. Il est de longueur variable, entre 600 et 500 résidus. Cependant, il est formé de 4 régions. La première région, nommée HRD, située juste en amont du domaine catalytique, est formé de répétitions HEAT de repliement similaire au domaine ARM des PI3Ks<sup>139</sup>. Cependant, le domaine ARM des PI3Ks est plus court que la région HRD. Or, si je superpose le modèle par homologie réalisé sur la région N-terminale de la PI4KA humaine à la structure cristallographique de la PIKK mTor (pdb 4JSV), le domaine ARM de la PI4KA correspond parfaitement à la longueur du domaine HRD. Plusieurs autres éléments confortent l'hypothèse que le domaine ARM correspond au sous-domaine HRD des PIKKs :

- Une mutation du résidu V1938, permettant l'interaction entre le N-lobe du domaine catalytique et le sous-domaine HRD, active mTor. Ceci est en accord avec l'hypothèse de phénomène d'allostérie impliquant les hélices H1 et H2 du domaine ARM de PI4KA
- En amont des hélices correspondant aux hélices H1 et H2 se trouvent deux hélices dont les extrémités possèdent des prolines.



**Figure 47- Comparaison des structures de différents membres de la superfamille des PI3K**

**En haut** : Représentation schématique des structures de mTor, PI4KA (PI4KIIIα), PI4KB (PI4KIIIβ), et PI3KC3

**En bas à gauche** : Structure cristallographique de l'enzyme-core de mTor, soit le module C-terminal compris entre TRD1 et CAT (code PDB : 4JSV)

**En bas à droite** : Modélisation des domaines ARM et CAT de la PI4KA humaine obtenue avec I-Tasser.

Pour chacune des structures situées au bas de la figure, le code couleur utilisé est le même que celui du schéma en haut de l'encart.

Le reste du domaine FAT est formé de trois autres sous-domaines : TRD1, 2 et 3. Il s'agit de sous-domaines hélicaux. Les sous-domaines TRD1 et HRD1 sont les segments les mieux conservés du domaine FAT chez les PIKKs, de même que DI est bien conservé chez les PI4KA/Stt4p. De plus, TRD1, situé en N-terminal de FAT, interagit directement avec le domaine catalytique des PIKKs. Des mutations à leur interface entraînent l'activation ou l'inactivation de la kinase. Le sous-domaine TRD1 équivaldrait donc au domaine putatif DI de PI4KA qui est indispensable à l'activité de la kinase de lipide. Ceci expliquerait aussi pourquoi les constructions DI ne sont pas solubles sans détergent puisque TRD1 n'est pas un domaine autonome. Par ailleurs, TRD2 et TRD3 correspondraient alors à DII et DIII. Cependant, TRD3 est entièrement hélical, à contrario de DIII qui serait plutôt désordonné. Comme pour DII-DIII la région TRD2-3 n'est pas très bien conservée.

Dans l'hypothèse où l'organisation de PI4KA DI-2102 serait semblable à celle de mTor TRD1-2459 (figure 47), on aurait une explication au fait que le domaine 1 de NS5A n'interagit pas avec les mutants de PI4KA déléts de DI<sup>28</sup> ni avec nos différentes constructions DI. En effet, ce domaine est essentiel dans le maintien de la structure de l'enzyme-core et donc dans la bonne conformation des surfaces d'interaction.

#### - Domaine Catalytique homologue aux PI3Ks

La purification du domaine catalytique de mTor a nécessité la présence du domaine FAT, sans lequel la protéine n'est pas active. La production a été réalisée en HEK-293F. La structure de la PIKK DNA-PKC à basse résolution a été obtenue sur la protéine entière produite en HeLa S3<sup>140</sup>. Ces deux souches sont des cellules de mammifères, attestant de la difficulté à produire de telles constructions. Cela peut expliquer en partie pourquoi la partie C-terminale de la PI4KA est aussi difficile à produire en bactérie et en cellule d'insecte.

Par ailleurs, la solubilité de mTor est fortement diminuée en l'absence de la protéine régulatrice mLST8. L'agrégation massive de la protéine lors de l'expression ainsi que le caractère hydrophobe de la PI4KA entière pourrait donc être dû aussi à l'absence de partenaire protéique. Ainsi, la construction insoluble 1129-2102 ainsi que la PI4KA entière sont certainement des constructions bien définies. Cependant, le système d'expression n'est pas adapté et il manque certainement un agent stabilisant de la protéine comme la protéine cellulaire TTC7 ou encore la présence de son substrat le phosphatidylinositol.

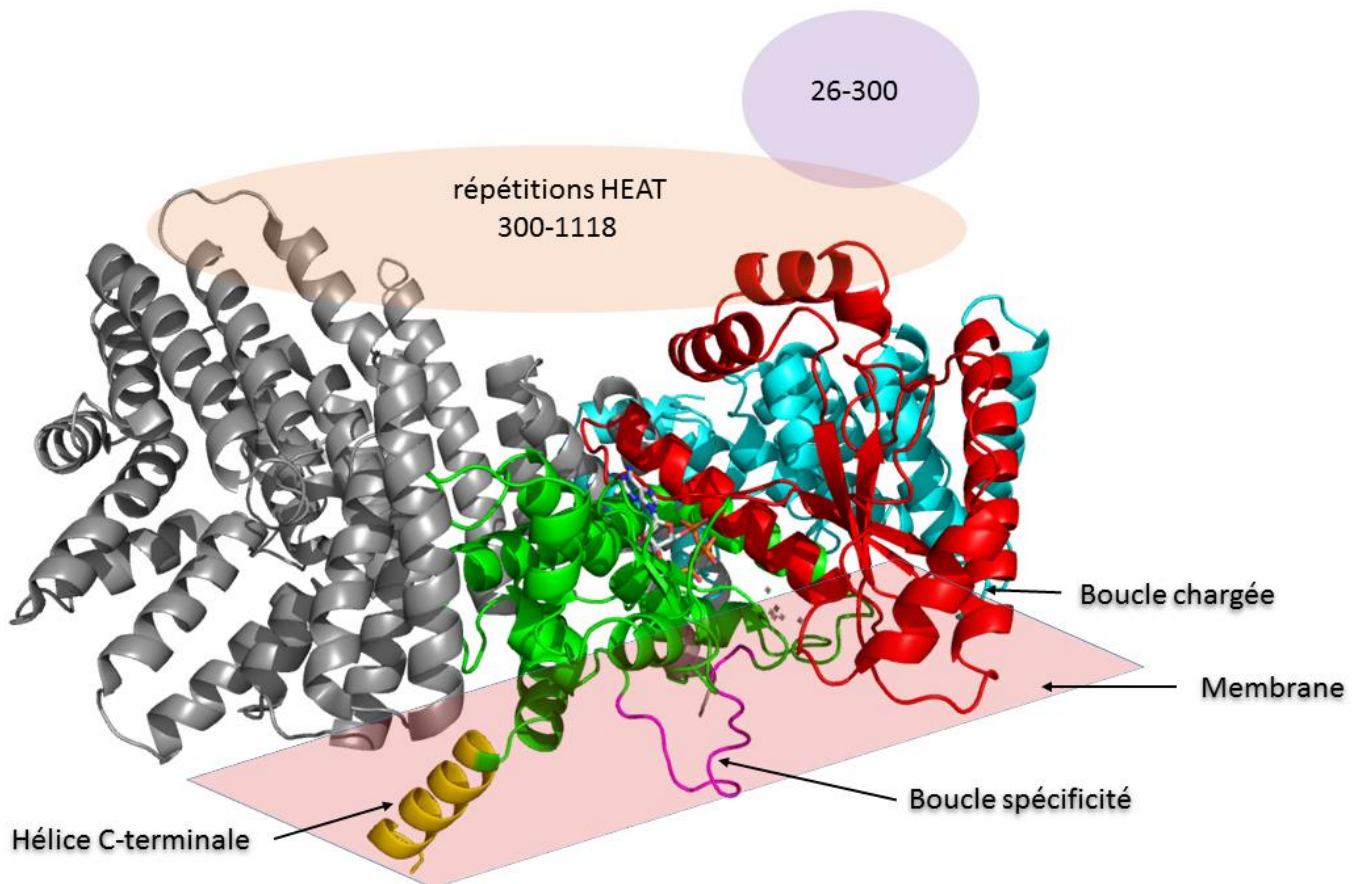
- **Domaine  $\alpha$ -hélicoïdal des PIKKs vs module N-terminal 26-1118**

Les PIKKs sont caractérisées par un grand domaine N-terminal  $\alpha$ -hélicoïdal de 1000 à 3000 résidus formant des super hélices. Ces domaines sont connus pour connaître de gros réarrangements lors d'interactions avec leurs nombreux partenaires<sup>79</sup>.

La fonction de ce domaine n'est pas bien définie, mais il interviendrait dans la régulation de l'activité kinase. Certains motifs conservés sont retrouvés entre mTor et PI4KA, en particulier le motif -LGL- appartenant au motif riche en leucine très conservée chez les PI4KAs/STT4p.

Si l'on tient compte des contraintes imposées par la chaîne polypeptidique, de l'orientation du domaine FAT de mTor ainsi que de l'orientation du site catalytique vers la membrane, une des conformations possible de l'enzyme-core est d'être en interaction directe avec la membrane suivant l'orientation présentée dans la figure 48. De ce fait, l'hélice C-terminale, la boucle chargée observée dans le domaine ARM, ainsi que la boucle de spécificité de PI4KA seraient orientés vers la membrane. De plus une boucle du lobe N-cat de la PI4KA riche en résidus basiques possède une tyrosine qui pointe alors en direction de la membrane. Enfin, le domaine DI est ainsi susceptible d'interagir aussi avec la membrane.

Au vu de la conformation du complexe de PP2A, qui est aussi un homologue de PIKK, nous pouvons imaginer que le module N-terminal vienne s'empiler au-dessus de l'enzyme-core, interagissant ainsi avec l'ARM et potentiellement avec l'insertion 1810-1840. Cette insertion pourrait ainsi, par le biais de phosphorylation, permettre en partie les changements de conformation de module N-terminal.



**Figure 48-** Proposition d'agencement des domaines et modules de la PI4KA humaine et d'une orientation possible par rapport à la membrane.

Dans cette conformation, l'hélice C-terminale, la boucle de spécificité (mauve), la boucle chargée de l'ARM interagissent directement avec la membrane.

Le modèle a été construit avec le modèle du domaine CAT-ARM en position ouverte minimisé avec AMBER. La structure cristallographique des domaines TRD1, 2 et 3 de mTor (PDB 4JSV) a été utilisée afin de signaler l'emplacement potentiel des domaines putatifs DI, DII et DIII de PI4KA. L'orientation de ce module a été déterminé par alignement du modèle de la partie C-terminale de PI4KA avec la structure entière de mTor.



**“Essentially, all models are wrong, but some are useful”**  
***(Box and Draper, 1987)***

L'étude de la PI4KA nous a mis dans une position typique où les informations sur la structure tri-dimensionnelle de la protéine sont pauvres et les études biophysiques difficilement réalisables. Dans ce cas précis, la modélisation computationnelle a été le moteur de nos recherches. Cependant, nous avons cherché au maximum des compléments expérimentaux.

Grâce à la communauté scientifique, un nombre croissant de familles de protéine possède au moins un membre pour lequel la structure à haute résolution est connue et repertoriée dans la PDB. Dans le même temps, la détection d'homologue par profil HMM s'est affinée permettant une meilleure exploitation de ces structures<sup>141</sup>. Malgré cela, la prédiction de la structure de protéine multi-domaine reste difficile à cause de leur forte plasticité.

Afin de tirer profit de la modélisation des protéines, nous avons choisi d'utiliser en majorité la modélisation par threading. Cette méthode est basée sur le principe que la structure est mieux conservée que la séquence primaire, permettant de meilleurs résultats que la modélisation par homologie lorsque la protéine d'intérêt possède moins de 25% d'identité ses homologues.

Globalement, cette stratégie nous a été utile puisque nous avons pu exploiter nos données SAXS grâce à la topologie prédite couplée à la recherche d'homologues lointains. D'autre part, cela m'a aussi permis de conforter la présence des domaines DI, DII et DIII. Cependant, même si les données étaient plutôt fiables, surtout pour le module N-terminal, il a été important de recouper le plus d'informations possibles entre elles afin de s'assurer de la fiabilité des données. Un exemple : j'ai pu me rendre compte que les modélisations générées par I-tasser étaient biaisées par un « effet de bord », d'où l'importance de séparer au maximum les domaines. Ainsi, la modélisation de toutes les constructions potentielles basées sur les prédictions de structure secondaire m'ont permis d'éliminer certaines des modélisations. Ce qui n'a pas empêché la mauvaise délimitation de certains domaines puisque expérimentalement le domaine 553-1118 semble être artéfactuel.

Par ailleurs, Yang Zhang lui-même explique que la plupart des méthodes de modélisation par threading ont du mal à détecter les meilleurs template quand la relation évolutive est faible. Voilà pourquoi, comme il est préconisé, j'ai décidé de séparer au maximum les domaines hypothétiques lors de mes analyses bio info par modélisation<sup>142</sup>.

Ainsi, l'utilisation du modèle doit être faite à bon escient et il est important de bien se rappeler que suivant la qualité du modèle, seule la structure globale, la topologie peuvent être utilisés alors que d'autres peuvent être utilisés un peu plus finement.

Même si le modèle de PIK4A entière que nous proposons est hypothétique, il permet d'apporter les bases nécessaires aux expérimentateurs : quelles mutations pourraient affecter l'activité de la protéine ? Quelles pourraient être les interrupteurs de cette activité ? Comment la protéine peut-elle interagir avec les membranes ? etc....

En conclusion, ce n'est pas le modèle en lui-même qui détermine à quel point il peut être utilisé, mais bien la personne qui l'exploite et l'interprète. Dans notre cas, les outils-bio-informatiques nous ont permis de comprendre à quoi pouvait correspondre notre enveloppe SAXS, la délimitation de domaines de PI4KA et enfin l'homologie avec les PIKKs qui pourtant n'a pas été détectée lors de nos recherches d'homologues lointain.



# Annexes

---



*Le domaine ARM étant très peu conservé entre PI3/4ks, seul le domaine CAT a été comparé aux PI3Ks.*

227

<i>PI4KA_chimpanze/1-2044</i>	247	CLSALSEVKVQPGCYLPSNPEAIVLDIDYKSGTP-MQSAAKAPYLAKFKV	295
<i>PI4KA_orang_outan/1-2019</i>	235	.....CYLPSNPEAIVLDIDYKSGTP-MQSAAKAPYLAKFKV	270
<i>PI4KA_humaine/1-2044</i>	247	CLSALSEVKVQPGCYLPSNPEAIVLDIDYKSGTP-MQSAAKAPYLAKFKV	295
<i>PI4KA_poulet/1-2123</i>	247	CLSALSEVKVQPGCYLPSNPEAIVLDIDYKSGTP-MQSAAKAPYLAKFKV	295
<i>PI4KA_Xenopus/1-2098</i>	247	CLNSLSEVKVQPGCYLPSNPEAIVLDIDYTSGETP-LQSAAKAPYLAKFKV	295
<i>PI4KA_zebrafish/1-2059</i>	249	CLKALSEIKVQPGCYLPSNPEAIVLDIDYKSGTP-MQSAAKAPYLAKFKV	297
<i>PI4KA_drosophile/1-2090</i>	243	CLACLKRIQVQPGCYLPSNPEAMVLDIDYNSGETP-MQSAAKAPYLARFRV	291
<i>Stt4p_AspERGillus/1-1803</i>	159	IEEELRKIKVEIGVYLPSPDGVVVIDIRKSGKP-LQSHAKAPYMATFRI	207
<i>Stt4p_S.Cerevisiae/1-1892</i>	228	IDDEMSKIEVKPDVYLPSPDGVVVIDIRKSGKP-LQSHAKAPYMATFKI	276
<i>PI4KB_drosophile/1-1066</i>	144	.....IDTSYLGEDWATKKERIRKTSEYGHFENWDL	174
<i>Pik1P_S.Cerevisiae/1-1066</i>	144	.....IDTSYLGEDWATKKERIRKTSEYGHFENWDL	174
<i>Vps34_humaine/1-887</i>	148	.....GIIPETATLTKSALMPAQLFFKT	170
<i>Vps34_S.Cerevisiae/1-875</i>	128	.....DVCPETSKVFKSSLSPLKITEKT	150
<i>PI3Kalpha_humaine/1-1068</i>	134	.....NLRLEECRIMSSAKRPLWLWNEN	156
<i>PI3Kdelta_humaine/1-1044</i>	134	.....EVCVEQCTFMDSKMKPLWIMYSN	156
<i>PI3Kgamma_humaine/1-1102</i>	141	.....ALAIKCKVMASKKKPLWLEFKC	163
<i>PI4KB_humaine/1-816</i>	126	.....WQEKVRRIRREGSPYGHLPNWR	147

Consensus

CLSALSE+KVQPGCYLPSNPEAIVLDIDYKSGTPRMQSAAKAPYLAKFKV

<i>PI4KA_chimpanze/1-2044</i>	296	KRCGVSELEKEGLRCRSDSEDECS.....TQEANGQKISWQAAI	334
<i>PI4KA_orang_outan/1-2019</i>	271	KRCGVSELEKEGLRCRSDSEDECS.....TQEADGQKISWQAAI	309
<i>PI4KA_humaine/1-2044</i>	296	KRCGVSELEKEGLRCRSDSEDECS.....TQEADGQKISWQAAI	334
<i>PI4KA_poulet/1-2123</i>	296	KRCGVSELEKEGLRCRSDSIDESA.....EEGVDSKKICWQAAI	334
<i>PI4KA_Xenopus/1-2098</i>	296	KRCGVSELEKEGLRCRSDSIDESA.....EEGDDG-KVCWQAAI	333
<i>PI4KA_zebrafish/1-2059</i>	298	KRCGVSELEKEGLRCRSDSQDEGD.....ESSLMAQRVCWQAAI	336
<i>PI4KA_drosophile/1-2090</i>	292	YRCGITETEFRAMEVSNPNPSQED.....AKMTLG-VESWQAAI	329
<i>Stt4p_AspERGillus/1-1803</i>	208	QKTRTRGDERESQGGSRSSSTAGHDASTKTLAPLHSPSSQETIEVWQSAI	257
<i>Stt4p_S.Cerevisiae/1-1892</i>	277	KKDVKDP.....LTGKNKEVEKWQAAI	298
<i>PI4KB_drosophile/1-1066</i>	175	C.....SVI	178
<i>Pik1P_S.Cerevisiae/1-1066</i>	175	C.....SVI	178
<i>Vps34_humaine/1-887</i>	171	EDGGK.....YPVI	179
<i>Vps34_S.Cerevisiae/1-875</i>	151	TLNQF.....YHLM	159
<i>PI3Kalpha_humaine/1-1068</i>	157	PDIMSELLFN.....NEI	171
<i>PI3Kdelta_humaine/1-1044</i>	157	EEAGSGGSGV.....I	168
<i>PI3Kgamma_humaine/1-1102</i>	164	ADPTALSNETIG.....I	177
<i>PI4KB_humaine/1-816</i>	148	LS.....VI	151

Consensus

KRCGVSELEKEGLRCRSDSEDEC+.....+Q+ADGQKISWQAAI

<i>PI4KA_chimpanze/1-2044</i>	335	FKVGDDCRQDMLALQIIDLFKNIFQLVGLDLFVFPYRVVATAPGCGVIEC	384
<i>PI4KA_orang_outan/1-2019</i>	310	FKVGDDCRQDMLALQIIDLFKNIFQLVGLDLFVFPYRVVATAPGCGVIEC	359
<i>PI4KA_humaine/1-2044</i>	335	FKVGDDCRQDMLALQIIDLFKNIFQLVGLDLFVFPYRVVATAPGCGVIEC	384
<i>PI4KA_poulet/1-2123</i>	335	FKVGDDCRQDMLALQIIDLFKNIFQLVGLDLFVFPYRVVATAPGCGVIEC	384
<i>PI4KA_Xenopus/1-2098</i>	334	FKVGDDCRQDMLALQIIGLFKNIFQLVGLDLFVFPYRVVATAPGCGVIEC	383
<i>PI4KA_zebrafish/1-2059</i>	337	FKVGDDCRQDMLALQIIGLFKNIFQLVGLDLFVFPYRVVATAPGCGVIEC	386
<i>PI4KA_drosophile/1-2090</i>	330	FKVGDDVRQDILALQVITIFKNIFQQVGLLEFLFPYRVVATAPGCGVIEC	379
<i>Stt4p_AspERGillus/1-1803</i>	258	FKVGDDCRQDMLALQMIIAAFRSIFASAGLDVWVFPYRVVATAPGCGVIDV	307
<i>Stt4p_S.Cerevisiae/1-1892</i>	299	FKVGDDCRQDVLALQLISLFRITWSSIGLDVYVFPYRVVATAPGCGVIDV	348
<i>PI4KB_drosophile/1-1066</i>	179	AKTGDDLRQEAFAAYQMIQAMANIWVKEKVDVWVKRMKILITSANTGLVET	228
<i>Pik1P_S.Cerevisiae/1-1066</i>	179	AKTGDDLRQEAFAAYQMIQAMANIWVKEKVDVWVKRMKILITSANTGLVET	228
<i>Vps34_humaine/1-887</i>	180	FKHGDDLRQDQLILQIISLMDKLLRKENLDLKLTPYKVLATSTKHGFQMF	229
<i>Vps34_S.Cerevisiae/1-875</i>	160	FKVGDDLRQDQLLVQIISLMNELLKNENVDLKLTPYKILATGPQEGAI	209
<i>PI3Kalpha_humaine/1-1068</i>	172	FKNGDDLRQDMLTLQIIRIMENIWQNGGLDLRMLPYGCLSIGDCTGLIEV	221
<i>PI3Kdelta_humaine/1-1044</i>	169	FKNGDDLRQDMLTLQIQLMDVWLKQEGDLRMTYPGCLPTGDRGLIEV	218
<i>PI3Kgamma_humaine/1-1102</i>	178	FKHGDDLRQDMLILQILRIMESIWETESLDLCLLPYGCISTGDKIGMIEI	227
<i>PI4KB_humaine/1-816</i>	152	VKCGDDLRQELLAFQVLKQLQSIWEQERVPLWIKPKILVISADSGMI	201

Consensus

FKVGDD+RQDMLALQIIDLFKNIFQL+GLDLFVFPYRV+ATAPGCGVIEC

<i>PI4KA_chimpanze/1-2044</i>	385	IPDCTSRDQLG.....RQTDGMYDYFTR.....QYGDDES	414
<i>PI4KA_orang_outan/1-2019</i>	360	IPDCTSRDQLG.....RQTDGMYDYFTR.....QYGDDES	389
<i>PI4KA_humaine/1-2044</i>	385	IPDCTSRDQLG.....RQTDGMYDYFTR.....QYGDDES	414
<i>PI4KA_poulet/1-2123</i>	385	IPDCTSRDQLG.....RQTDGMYDYFTR.....QYGDDES	414
<i>PI4KA_Xenopus/1-2098</i>	384	IPDCTSRDQLG.....RQTDGMYDYFTR.....QYGDDES	413
<i>PI4KA_zebrafish/1-2059</i>	387	IPDCKSRDQLG.....RQTDGMYDYFRN.....QYGDDES	416
<i>PI4KA_drosophile/1-2090</i>	380	VPNAKSRDQLG.....RQTDGLEYEFLH.....QYGDDES	409
<i>Stt4p_AspERGillus/1-1803</i>	308	LPNSISRDMLG.....REAVNGLYDYFIS.....KYGGEH	337
<i>Stt4p_S.Cerevisiae/1-1892</i>	340	LPNSVSRDMLG.....REAVNGLEYEFTS.....KFGNES	378
<i>PI4KB_drosophile/1-1066</i>	229	ITNAMSIVHSIKKALTKKMIEDAELDDKGGIASLNDHFLR.....AFGNPN	273
<i>Pik1P_S.Cerevisiae/1-1066</i>	229	ITNAMSIVHSIKKALTKKMIEDAELDDKGGIASLNDHFLR.....AFGNPN	273
<i>Vps34_humaine/1-887</i>	230	IQS-VPVAEVLDT.....EGSIQNFRRKYAP.....SENGP	259
<i>Vps34_S.Cerevisiae/1-875</i>	210	IPN-DTLASILSK.....YHGILGYLKLHYP.....DENAT	239
<i>PI3Kalpha_humaine/1-1068</i>	222	VRNSHTIMQIQ.....KGGKGLALQFNSHTLHQLWKDKNK	257
<i>PI3Kdelta_humaine/1-1044</i>	219	VLRSDTIANIQLN.....KSNMAATAAFNKDALLNWLKSKNP	255
<i>PI3Kgamma_humaine/1-1102</i>	228	VKDATTIAKIQS.....TVGNTGAFKDEVLN-HWLKESPT	263
<i>PI4KB_humaine/1-816</i>	202	VVNAVSIHQVK.....KQSLSLLDYFLQ.....EHGYSY	231

Consensus

IPNCTSRDQLGKALTKKMIEDAELDDRQTDG+YDYFTR-LH-WQYGDDES



*PI4KA\_chimpanze/1-2044* 415 TLA FQQ ARY NF I RSMAAYS LLL F LLQ IKDRHNGNIMLDKKGH I I H I D F G F 464  
*PI4KA\_orang\_outan/1-2019* 390 TLA FQQ ARY NF I RSMAAYS LLL F LLQ IKDRHNGNIMLDKKGH I I H I D F G F 439  
*PI4KA\_humaine/1-2044* 415 TLA FQQ ARY NF I RSMAAYS LLL F LLQ IKDRHNGNIMLDKKGH I I H I D F G F 464  
*PI4KA\_poulet/1-2123* 415 TLA FQQ ARY NF I RSMAAYS LLL F LLQ IKDRHNGNIMLDKKGH I I H I D F G F 464  
*PI4KA\_Xenopus/1-2098* 414 ALA FQK ARY NF I RSMAAYS LLL F LLQ IKDRHNGNIMLDKKGH I I H I D F G F 463  
*PI4KA\_zebrafish/1-2059* 417 TLA FQK ARY NF I RSMAAYS LLL F LLQ IKDRHNGNIMLDKKGH I I H I D F G F 466  
*PI4KA\_drosophile/1-2090* 410 SKEFQAARANFVKSMAAYS L I G Y LLQ IKDRHNGNIMIDKDGHV I H I D F G F 469  
*Stt4p\_Aspgillus/1-1803* 338 SMRFQEARTNFVKSMAAYS V I S Y LLQ F KDRHNGNIMIDDA G H I I H I D F G F 387  
*Stt4p\_S.Cerevisiae/1-1892* 379 T I E F Q N ARN NF V K S L A G Y S V I S Y LLQ F KDRHNGNIMYD D Q G H C L H I D F G F 428  
*PI4KB\_drosophile/1-1066* 274 G F K Y R R A Q D N F A S S L A A Y S V I C Y LLQ V KDRHNGNIMIDNE G H V S H I D F G F 323  
*Pik1P\_S.Cerevisiae/1-1066* 274 G F K Y R R A Q D N F A S S L A A Y S V I C Y LLQ V KDRHNGNIMIDNE G H V S H I D F G F 323  
*Vps34\_humaine/1-887* 260 N G I S A E V M D T Y V K S C A G Y C V I T Y I L G V G D R H L D N L L L T K T G K L F H I D F G Y 309  
*Vps34\_S.Cerevisiae/1-875* 240 L G V Q G W V L D N F V K S C A G Y C V I T Y I L G V G D R H L D N L L V T P D G H F F H A D F G Y 289  
*PI3Kalpha\_humaine/1-1068* 258 G E I Y D A A I D L F T R S C A G Y C V A T F I L G I G D R H N S N I M V K D D G Q L F H I D F G H 307  
*PI3Kdelta\_humaine/1-1044* 256 G E A L D R A I E E F T L S C A G Y C V A T V L G I G D R H S D N I M I R E S G Q L F H I D F G H 305  
*PI3Kgamma\_humaine/1-1102* 264 E E K F Q A A V E R F V S C A G Y C V A T F V L G I G D R H N D N I M I T E T G N L F H I D F G H 313  
*PI4KB\_humaine/1-816* 232 T E A F L S A Q R N F V Q S C A G Y C L V C Y LLQ V KDRHNGN I L L D A E G H I I H I D F G F 281

Consensus



*PI4KA\_chimpanze/1-2044* 465 M F E S S P G G N L G W E P D I K L T D E M V M I M G G K M E . . . . . 495  
*PI4KA\_orang\_outan/1-2019* 440 M F E S S P G G N L G W E P D I K L T D E M V M I M G G K M E . . . . . 470  
*PI4KA\_humaine/1-2044* 465 M F E S S P G G N L G W E P D I K L T D E M V M I M G G K M E . . . . . 495  
*PI4KA\_poulet/1-2123* 465 M F E S S P G G N L G W E P D I K L T D E M V M I M G G K M E . . . . . 495  
*PI4KA\_Xenopus/1-2098* 464 M F E S S P G G N L G W E P D I K L T D E M V M I M G G K M E . . . . . 494  
*PI4KA\_zebrafish/1-2059* 467 M F E S S P G G N L G W E P D I K L T D E M V M I M G G K M E . . . . . 497  
*PI4KA\_drosophile/1-2090* 460 M F E S S P G G N I G F E P D M K L T D E M V M I M G G K M E . . . . . 490  
*Stt4p\_Aspgillus/1-1803* 388 C F D I A P G G V R F E R A P F K L T S E M V A V M S G T S H P H I H S H S G S N G G L G L P G G H 437  
*Stt4p\_S.Cerevisiae/1-1892* 429 I F D I V P G G I K F E A V P F K L T K E M V K V M G G S P Q . . . . . 469  
*PI4KB\_drosophile/1-1066* 324 M L S N S P G S V G F E A A P F K L T Y E Y I E L L G G I E G . . . . . 354  
*Pik1P\_S.Cerevisiae/1-1066* 324 M L S N S P G S V G F E A A P F K L T Y E Y I E L L G G I E G . . . . . 354  
*Vps34\_humaine/1-887* 310 I L G R . . . . . D P K P L P P M K L N K E M V E G M G . . . . . 333  
*Vps34\_S.Cerevisiae/1-875* 290 I L G Q . . . . . D P K P F P P L M K L P P Q I I E A F G . . . . . 313  
*PI3Kalpha\_humaine/1-1068* 308 F L D H K K K K F G Y K R E R V P F V L T Q D F L I V I S K G A Q . . . . . 340  
*PI3Kdelta\_humaine/1-1044* 306 F L G N F K T K F G I N R E R V P F I L T Y D F V H V I Q Q G K T . . . . . 338  
*PI3Kgamma\_humaine/1-1102* 314 I L G N Y K S F L G I N K E R V P F V L T P D F L F V M G T S G K . . . . . 346  
*PI4KB\_humaine/1-816* 282 I L S S S P R N L G F E T S A F K L T T E F V D V M G G L D G . . . . . 312

Consensus



*PI4KA\_chimpanze/1-2044* 496 . . . . . A T P F K W F M E M C V R G Y L A V R P Y M D A V V S L V T L M L D T - G L P C F R . . 536  
*PI4KA\_orang\_outan/1-2019* 471 . . . . . A T P F K W F M E M C V R G Y L A V R P Y M D A V V S L V T L M L D T - G L P C F R . . 511  
*PI4KA\_humaine/1-2044* 496 . . . . . A T P F K W F M E M C V R G Y L A V R P Y M D A V V S L V T L M L D T - G L P C F R . . 536  
*PI4KA\_poulet/1-2123* 496 . . . . . A T P F K W F M E M C V R G Y L A V R P Y M D A V V S L V T L M L D T - G L P C F R . . 536  
*PI4KA\_Xenopus/1-2098* 495 . . . . . A T P F K W F M E M C V R G Y L A V R P Y M D A V V S L V T L M L D T - G L P C F R . . 535  
*PI4KA\_zebrafish/1-2059* 498 . . . . . A T P F K W F M E M C V R G Y L A V R P Y M D A V V S L V T L M L D T - G L P C F R . . 538  
*PI4KA\_drosophile/1-2090* 491 . . . . . A P A F K W F C E L C V Q A F L A V R P Y Q D A I V S L V S L M L D T - G L P C F R . . 531  
*Stt4p\_Aspgillus/1-1803* 438 S N N P T S T Q P Y R W F E S L V I K A F L A S R P F C T K L S H I V S L M L D S - G L P C F K . . 484  
*Stt4p\_S.Cerevisiae/1-1892* 460 . . . . . T P A Y L D F E E L C I K A Y L A A R P H V E A I I E C V N P M L G S - G L P C F K G - 501  
*PI4KB\_drosophile/1-1066* 355 . . . . . E A F K K F V E L T K S S F K A L R K Y A D Q I V S M C E I M Q K D N M Q P C F D A G 397  
*Pik1P\_S.Cerevisiae/1-1066* 355 . . . . . E A F K K F V E L T K S S F K A L R K Y A D Q I V S M C E I M Q K D N M Q P C F D A G 397  
*Vps34\_humaine/1-887* 334 . . . G T Q S E Q Y Q E F R K Q C Y T A F L H L R R Y S N L I L N L F S L M V D A N I P D I A L E P 380  
*Vps34\_S.Cerevisiae/1-875* 314 . . . G A E S S N Y D K F R S Y C F V A Y S I L R R N A G L I L N L F E L M K T S N I P D I R I D P 360  
*PI3Kalpha\_humaine/1-1068* 341 . . . E C T K T R E F E R F G E M C Y K A Y L A I R Q H A N L F I N L F S M M L G S G M P E L Q S F D 388  
*PI3Kdelta\_humaine/1-1044* 339 . . . N N S E K . . F E R F G Y C E R A Y T I L R H G L L F L H L F A L M R A A G L P E L S C S K 384  
*PI3Kgamma\_humaine/1-1102* 347 . . . . . K T S P H F Q K F Q D I C V K A Y L A L R H H T N L L I I L F S M M L M T G M P Q L T S . . 390  
*PI4KB\_humaine/1-816* 313 . . . . . D M F N Y Y K M L M L Q G L I A A R K H M D K V V Q I V E I M Q Q G S Q L P C F H G . 354

Consensus



*PI4KA\_chimpanze/1-2044* 537 G Q T I K L L K H R F S P N M T E R E A A N F I M K V I Q - S C F L S N R S R T Y D M I Q Y Y Q N D 585  
*PI4KA\_orang\_outan/1-2019* 512 G Q T I K L L K H R F S P N M T E R E A A N F I M K V I Q - S C F L S N R S R T Y D M I Q Y Y Q N D 580  
*PI4KA\_humaine/1-2044* 537 G Q T I K L L K H R F S P N M T E R E A A N F I M K V I Q - S C F L S N R S R T Y D M I Q Y Y Q N D 585  
*PI4KA\_poulet/1-2123* 537 G Q T I K L L K H R F S P N M T E R E A A T F I I K I I Q - N C F L S N R S R T Y D M I Q Y Y Q N D 585  
*PI4KA\_Xenopus/1-2098* 536 G Q T I K L L K Q R F S P N S T E R E A A N F I L K V I Q - N C F L S N R S R T Y D M I Q Y Y Q N D 584  
*PI4KA\_zebrafish/1-2059* 539 G Q T I K L L K Q R F S P N M S E K E A A A F M I K V I E - R C F L S S R S K T Y D M L Q Y Y Q N Q 587  
*PI4KA\_drosophile/1-2090* 532 G Q T I N L L K Q R F V A T K N N K E A A H M L A V I R - N S Y Q N F R T R T Y D M I Q Y Y Q N Q 580  
*Stt4p\_Aspgillus/1-1803* 485 P D T L K N F R D R F V L E K N E R E A A E Y M R E L I R - K S Y M S V S T K G Y D Q F Q L L T N G 533  
*Stt4p\_S.Cerevisiae/1-1892* 502 H K T I R N L R A R F Q P Q K T D H E A A L Y M K A L I R - K S Y E S I F T K G Y D E F Q . . . . 545  
*PI4KB\_drosophile/1-1066* 398 E Q T S V Q L R Q R F Q L D L S E K E V D D F V E N F L I G K S L G S I Y T R I Y D Q F Q L I T Q G 447  
*Pik1P\_S.Cerevisiae/1-1066* 398 E Q T S V Q L R Q R F Q L D L S E K E V D D F V E N F L I G K S L G S I Y T R I Y D Q F Q L I T Q G 447  
*Vps34\_humaine/1-887* 381 D K T V K K V Q D K F R L D L S D E E A V H Y M Q S L I D - E S V H A L F A A V V E Q I H K F A Q Y 429  
*Vps34\_S.Cerevisiae/1-875* 361 N G A I L R V R E R F N L N M S E E D A T V H F Q N L I N - D S V N A L L P I V I D H L H N L A Q Y 409  
*PI3Kalpha\_humaine/1-1068* 389 D . . I A Y I R K T L A L D K T E Q E A L E Y F M K Q M N - D A H H G G W T T K M D W I F H T I K Q 435  
*PI3Kdelta\_humaine/1-1044* 385 D . . I Q Y L K D S A L G K T E E A L K H F R V K F N - E A L R E S W K T K V N W L A H N V S K 431  
*PI3Kgamma\_humaine/1-1102* 391 K E D I E Y I R D A L T V G K N E E A K K Y F L D Q I E - V C R D K G W T V Q F N W F L H L V L G 439  
*PI4KB\_humaine/1-816* 355 S S T I R N L K E R F H M S M T E E Q L Q L L V E Q M V D - G S M R S I T T K L Y D G F Q Y L T N G 403

Consensus



<i>PI4KA_chimpanze/1-2044</i>	586	I P Y - . . . . .	588
<i>PI4KA_orang_outan/1-2019</i>	561	I P Y - . . . . .	563
<i>PI4KA_humaine/1-2044</i>	586	I P Y - . . . . .	588
<i>PI4KA_poulet/1-2123</i>	586	I P Y - . . . . .	588
<i>PI4KA_Xenopus/1-2098</i>	585	I P Y - . . . . .	587
<i>PI4KA_zebrafish/1-2059</i>	588	I P Y - . . . . .	590
<i>PI4KA_drosophile/1-2090</i>	581	I P Y - . . . . .	583
<i>Stt4p_AspERGILLUS/1-1803</i>	534	I P Y - . . . . .	536
<i>Stt4p_SCerevisiae/1-1892</i>	-	- . . . . .	
<i>PI4KB_drosophile/1-1066</i>	448	I Y S - . . . . .	450
<i>Pik1P_S.Cerevisiae/1-1066</i>	448	I Y S - . . . . .	450
<i>Vps34_humaine/1-887</i>	430	W R K - . . . . .	432
<i>Vps34_S.Cerevisiae/1-875</i>	410	W R T - . . . . .	412
<i>PI3Kalpha_humaine/1-1068</i>	436	H A L N - . . . .	439
<i>PI3Kdelta_humaine/1-1044</i>	432	D N R Q - . . . .	435
<i>PI3Kgamma_humaine/1-1102</i>	440	I K Q G E K H S A	448
<i>PI4KB_humaine/1-816</i>	404	I M - . . . . .	405

Consensus

I P Y - . . . . .

## ANNEXE 2 : ALIGNEMENT DE SEQUENCE MODULE DI-DII

DI

humain	1134	KDYSNFMASLNLNRYAG	.....	EVYGMIRFSGTTGQ	.....	MSDLNKMVMQDLHSLDRSHPHQYTGAMFKLT	AMLI	SSKDCDPQLLHHL	1215
chimpanze	1076	KDYSNFMASLNLNRYAG	.....	EVYGMIRFSGTTGQ	.....	MSDLNKMVMQDLHSLDRSHPHQYTGAMFKLT	AMLI	SSKDCDPQLLHHL	1157
Orang	1076	KDYSNFMASLNLNRYAG	.....	EVYGMIRFSGTTGQ	.....	MSDLNKMVMQDLHSLDRSHPHQYTGAMFKLT	AMLI	SSKDCDPQLLHHL	1157
Poule	1155	KDYSNFMASLNLNRYAG	.....	EVSGMIQFSNATGR	.....	TSDLNKLMLVKELNNALESQGTEFYTGAMFKLT	ALLI	SSKDCDPQLLHHL	1236
Marsupiale_Monodelphis	5	KDYSNFMASLNLNRYAG	.....	EVYGMIRFSDSTGH	.....	TSDLKKMLIKQLTSALDSQTEFYTGAMFKLT	AMLI	SLPECDPQLLHHL	86
Xenopus	1082	KDYSNFMASLNLNRYAG	.....	EVAGMIEFSDATGR	.....	TSDLNRLMLKRLSDALDCGPEAYTGVMFKLT	ALLI	SSKECDTOLLHHL	1163
Zebrafish	1076	KDYSNFMASLNLNRYAG	.....	EVAGMIHFSEATR	.....	TSDLNKLMLIKRQNEALDGGQPEAFTEALFKMAALLI	SSREWDPLLHHL	1157	
Poisson_Tetraodon	5	KDYSNFMASLNLNRYAG	.....	EVSGMIHFAQAVRG	.....	LSDLSQLMIKQMGAEALDSGQPEAFTEALFKMAALLI	SSKECDTOLLHHL	86	
Drosophila	5	SDTPRFVSVLCLRSKYAG	.....	EISGLLSVLSEQ	.....	DKAGLAERLVSDVWEACRDKSDARHRRGALWRAT	AYLI	ICAEVDRKLLHVA	85
moustique	5	SDTSKVFVSVLCLRSKYAG	.....	EISGLLSVLEDE	.....	EKRGLADRLVREIWAACEEKSDARHRRGALWRAT	AYLI	ICSSANRKLHAI	85
guepe_Nasonia	4	DANSWLLSMMSTRSRVAG	.....	EVAGMLNLVrgesptesDSM	.....	KALVNRLVEEILRACRNRDDNEHRRALWRAT	ALLI	SIPGLNRKLLHTVA	90
bourdon	5	GASSWLLSMMSTRSRVAG	.....	EIAGMLSLaqtesstseISFEEIRNKVIDLL	.....	DAVWMACRARNRVNHRALWRAT	ALLI	SMGPOTHRRLHTVA	95
Foumi_Harpegnathos	5	GASSWLLSMMSTRSRVAG	.....	EVAGMLALartespaeMSLEESRDQKVDVRI	.....	IEAVHRACQRNDAAHRRGALWRAT	ALLI	SMGPOTHRRLHTVA	95
coleoptere_Tribolium	5	SDSSKLMVSTSLRCKYIG	.....	EIVGLLSAVYGT	.....	GKEKLIIEFIMKRVWTAQCRSDSEHRDALWQAT	ALLI	STTEFHRNLLHCA	85
tique_kodes	5	NNCSEFVCSMLRCRFAG	.....	EITGLLMGaeidi	.....	DAMHKLSSRLNLQKLSWEKKDEHIHKECIFRVC	ALLI	HKSGDTRALLHVA	87
pou	5	SDSSRFITSSQLRSRYMG	.....	EVMLLMNEG	.....	EKHKLITQLVDKVVESCAMKDDETHGAALWRAT	ALLI	LPMDRRLHVA	82
crustace_Daphnia	5	CDSPRFIAVLGKNRYSAG	.....	QVAGMFSCLQsdTa	.....	vtkdKQVEKLSAQLMEDMRQSLAKDEKAHCRALWCIA	ALVIRIDGVCNRLHSA	92	
nematode_Caenorhabditis	1	SGSYLTSLYkrnsyig	.....	qvkmgiaattlsgdi	.....	dqVTAENKIIDREKDFQRACKSGKDNEMEQAVLLLT	ALFVSLRCLNERILTLLV	85	
nematode_Ascaris	3	PaasIYLSLhrlslyig	.....	qvkmgiamckeds	.....	TKAELRIIEHLECDLSRACETGCEEMTDAVMKLAALFI	SIKELNTRLLHTLV	87	
cerevisiae	1022	NTISGIFSLHSWRSKLYLFDTAITSSPEDIKRQIGIST	.....	.....	.....	QNIKNTLLGNKIITKDVDTFLDMAT	ALLI	LGNGAPASLYIDV	1102
neurospora	1018	NTASDFMAQYTKRQYRYTEAVPDRSL	.....	EWQGMHDPQATSSLPVTS	.....	SDSLDVTALAHIEKRVLSRKTTP	LDVRNTRLRRAA	ILCRSTDDECSI	1117
aspergillus	963	NVASDFIAQYTIKQYRAPDPLATKYLE	.....	.....	.....	.....	.....	QGSEKSKAAAILCGSDYFQPSILHHLV	1017

Consensus



humain	1216	WGFLRMFNEHGMETALAC	EWLLAGKDGVEVP	MRMAGAWHMTVEQKGLFSAEIK	.....	EADPLAASEASQP	.....	KPCPPEVTPHYI	1303
chimpanze	1158	WGFLRMFNEHGMETALAC	EWLLAGKDGVEVP	MRMAGAWHMTVEQKGLFSAEIK	.....	EADPLAASEASQP	.....	KPCPPEVTPHYI	1245
Orang	1158	WGFLRMFNEHGMETALAC	EWLLAGKDGVEVP	MRMAGAWHMTVEQKGLFSAEIK	.....	EADPLAASEASQP	.....	KPCPPEVTPHYI	1245
Poule	1237	WGFLRMFNEHGMETALAC	EWLLAGKDGVEVP	MRMAGAWHMTVEQKGLFSAEIK	.....	EADPLAASEASQP	.....	KPCPPEVTPHYI	1324
Marsupiale_Monodelphis	87	WGFLRMFNEHGMETALAC	EWLLAGKDGVEVP	MRMAGAWHMTVEQKGLFSAEIK	.....	EADPLAASEASQP	.....	KPCPPEVTPHYI	174
Xenopus	1164	WAPLRFNEHGMETALAC	EWLLAGKDGVEVP	MRMAGAWHMTVEQKGLFSAEIK	.....	EADPLAASEASQP	.....	KPCPPEVTPHYI	1251
Zebrafish	1158	WAPLRFNEHGMETALAC	EWLLAGKDGVEVP	MRMAGAWHMTVEQKGLFSAEIK	.....	EADPLAASEASQP	.....	KPCPPEVTPHYI	1245
Poisson_Tetraodon	87	WGFLRMFNEHGMETALAC	EWLLAGKDGVEVP	MRMAGAWHMTVEQKGLFSAEIK	.....	EADPLAASEASQP	.....	KPCPPEVTPHYI	174
Drosophila	86	SSQLELFTSVMETAVECQWVLTARQDL	ELCFIQEMVSAWQTTFEKRMGLFALEQE	.....	.....	VTPLAAYEGCKL	.....	VSSPILIVPHL	173
moustique	86	SSQVLLFTSVMETAVECQWVLTARQDL	ELCFIQEMVSAWQTTFEKRMGLFALEQE	.....	.....	VTPLAAYEGCKL	.....	VSSPILIVPHL	173
guepe_Nasonia	91	CSQLELFTKPMRTAVECQWVLTARQDL	ELCFIQEMVSAWQTTFEKRMGLFALEQE	.....	.....	VTPLAAYEGCKL	.....	VSSPILIVPHL	178
bourdon	96	SSQLELFTKPMRTAVECQWVLTARQDL	ELCFIQEMVSAWQTTFEKRMGLFALEQE	.....	.....	VTPLAAYEGCKL	.....	VSSPILIVPHL	178
Foumi_Harpegnathos	96	SSQLELFTKPMRTAVECQWVLTARQDL	ELCFIQEMVSAWQTTFEKRMGLFALEQE	.....	.....	VTPLAAYEGCKL	.....	VSSPILIVPHL	178
coleoptere_Tribolium	86	SSQVLLFTSVMETAVECQWVLTARQDL	ELCFIQEMVSAWQTTFEKRMGLFALEQE	.....	.....	VTPLAAYEGCKL	.....	VSSPILIVPHL	173
tique_kodes	86	SSQVLLFTSVMETAVECQWVLTARQDL	ELCFIQEMVSAWQTTFEKRMGLFALEQE	.....	.....	VTPLAAYEGCKL	.....	VSSPILIVPHL	173
pou	83	WSQVLEFTSVMETAVECQWVLTARQDL	ELCFIQEMVSAWQTTFEKRMGLFALEQE	.....	.....	VTPLAAYEGCKL	.....	VSSPILIVPHL	170
crustace_Daphnia	93	WSHLELFTSVMETAVECQWVLTARQDL	ELCFIQEMVSAWQTTFEKRMGLFALEQE	.....	.....	VTPLAAYEGCKL	.....	VSSPILIVPHL	182
nematode_Caenorhabditis	86	RTPLHNFTESTISLCTMAWNL	SARPEAHSFLHEqARAFADSCRFGLG	.....	.....	FEKENVdMSISPLCRQV	.....	PERP	173
nematode_Ascaris	86	SVPLRFNTASTMRLCVSNWNI	LVARDIHLSELRemaSCW	.....	.....	INVAQKAGLfeIDAP	.....	FvSPLSVQY	173
cerevisiae	1103	HIPEVFNTSASLKTVMNWLT	ITEKPEVAHLLVVCYCWMRS	.....	.....	IDDNIGLYSRDHD	.....	LKGEYQMKMEYSPYDKAG	1200
neurospora	1118	SIPFGIFTKQSIKGLVSLWGLV	INENPRMEPRILAEIAQQWEFS	.....	.....	IQKRLGLFSSTIC	.....	NPDPFFLKEEFAPTDSASVAKRKL	1215
aspergillus	1018	ILPFOLFSEFSMNVGVALWGA	IHENPKLGPRILAEIAWEET	.....	.....	IQRRKGLFDRSFD	.....	YVDPMTYKIELLP	1115

Consensus



humain	1304	QRFEIAKYCSSDQVEIFSSLLQRS	MSLNI	GG	.....	AKGSMNRHVAAIGPRFKLL	TLGLSLLHADVVP	NATIRNV	1381
chimpanze	1248	QRFEIAKYCSSDQVEIFSSLLQRS	MSLNI	GG	.....	AKGSMNRHVAAIGPRFKLL	TLGLSLLHADVVP	NATIRNV	1323
Orang	1248	QRFEIAKYCSSDQVEIFSSLLQRS	MSLNI	GG	.....	AKGSMNRHVAAIGPRFKLL	TLGLSLLHADVVP	NATIRNV	1323
Poule	1325	QRFEIAKYCSSDQVEIFSSLLQRS	MSLNI	GG	.....	AKGSMNRHVAAIGPRFKLL	TLGLSLLHADVVP	NATIRNV	1402
Marsupiale_Monodelphis	175	QRFEIAKYCSSDQVEIFSSLLQRS	MSLNI	GG	.....	AKGSMNRHVAAIGPRFKLL	TLGLSLLHADVVP	NATIRNV	252
Xenopus	1252	QRFEIAKYCSSDQVEIFSSLLQRS	MSLNI	GG	.....	AKGSMNRHVAAIGPRFKLL	TLGLSLLHADVVP	NATIRNV	1329
Zebrafish	1248	QRFEIAKYCSSDQVEIFSSLLQRS	MSLNI	GG	.....	AKGSMNRHVAAIGPRFKLL	TLGLSLLHADVVP	NATIRNV	1323
Poisson_Tetraodon	175	QRFEIAKYCSSDQVEIFSSLLQRS	MSLNI	GG	.....	AKGSMNRHVAAIGPRFKLL	TLGLSLLHADVVP	NATIRNV	252
Drosophila	174	EMVDTAKYCNDRKQVEMFCL	LLHRC	LPILKSS	.....	.....	.....	.....	248
moustique	174	EMVDTAKYCNDRKQVEMFCL	LLHRC	LPILKSS	.....	.....	.....	.....	248
guepe_Nasonia	179	DLIETAKYCCQESVEMVALL	LLHRC	LPILKSS	.....	.....	.....	.....	259
bourdon	184	ELIETAKYCCQETVDMIVT	LLHRS	LPMTV	.....	.....	.....	.....	263
Foumi_Harpegnathos	184	ELIETAKYCCQETVDMIVT	LLHRS	LPMTV	.....	.....	.....	.....	263
coleoptere_Tribolium	174	ELIDNVKYSSYEKVEMLAS	ILHRS	LPMTV	.....	.....	.....	.....	251
tique_kodes	176	ELIETAKYSSLDQVEIFAH	ILHRS	LPMTV	.....	.....	.....	.....	253
pou	171	ELIETAKYSSLDQVEIFAH	ILHRS	LPMTV	.....	.....	.....	.....	248
crustace_Daphnia	183	ERIEIAKHCSQDQVELFAQ	MFDTLP	PFdVgS	.....	.....	.....	.....	260
nematode_Caenorhabditis	174	ERCDVAYSSEqLDv	lemmfsgLSAQIGT	ghgstaonts	.....	.....	.....	.....	271
nematode_Ascaris	174	ERCDVAYSSEqLDv	lemmfsgLSAQIGT	ghgstaonts	.....	.....	.....	.....	271
cerevisiae	1201	SHFEGTLFQSDFLKIF	FTKCALY	GLKNI	.....	.....	.....	.....	1273
neurospora	1216	SHYNATRLGNPDTRQIF	FLRL	LEVLT	.....	.....	.....	.....	1288
aspergillus	1116	SHFNVAURLNSAQNRRL	FSRL	IDRTT	.....	.....	.....	.....	1188

Consensus



DII





# ANNEXE 3 : ALIGNEMENT DE SEQUENCES DU MODULE N-TERMINAL DES PI4KA/STT4P

humain	1	MAAAPARGGGGGGGGGGGSGSGSSASRGFYFNTVLSLARSLAVQRPASLEKVVQKLLCMCP	81	
chimpanze	1		MCP 3	
Orang	1		MCP 3	
Poule	1	MAAATVAAAAAAT - - - AAAAAATAAAPTAAAAAPSAGVGRGSSAGTARGFYFNTVLSLARSLAVQSPAPLEKVVQKLLCMCP	82	
Xenopus	1		MCP 3	
poisson_Tetraodon	1		MCP 3	
drosophile_Drosophila	1	MTMTASDKYTYQRAVLCLARVLAGIQPTPWEKVQTLFRYCP	41	
Zebrafish	1		MCP 3	
guepe	1	MAADEKLIFSKIVQHLARSLAAIKNTPWEKVNWLKRC	39	
moustique	1	MVGNERFSFQKTQVCLARVLAMIKPTPWEKVQTLFRYCP	39	
foumi_Haepgnathos	1	MMADEKLFFSNAVQQLARCLAAIKPTPWEKVAKLYKLC	39	
pou	1	MSLNEKLFFRAIQHLARSLAAISKTPWEKVEKLLFC	39	
tique_kodes	1	MAIACSDRKFFCQAVQHLARSLACMQPTPPSKLDSLRLC	41	
crustace_Daphnia	1	MIETSDFFSKTVQHLARSLAHIENTPWEKVSNLFLNLC	38	
nematode_Ascaris	1	MAIDEDFLQNLQCTAICLARSNKVTVEQVKKLLLENV	38	
nematode_Caenorhabditis	1	MRDIEDFVFNTKQCLAVCIKNNENTKLSQIFQLLALDD	38	
cerevisiae	1	MRFTRGLK - - - - - ASSSLRAKAI GRLTKLSTG - - - APNDQNSNGTLDLITHTLPIFYSTNTSKLYT - - - IPLTLSEWEVLTSCLV	75	
aspergillus	1		MKGSRELDVLLALCD	15
neurospora	1	MGRNIRSKALLQIAALSASNPAASFQSDLDKLCRVSVTQGHSAPEGAVNGRGP - - - SSPSSVPPGGPNVGR - - - IPMTIKEYEVLFA	84	

Consensus

---R---A---A---AAAS---L---SAA---APGA-N+-G---+SMMA+E++F++VQ+LARSLAAIKPTPWEKVQKLL++CP

humain	62	VDF--HGIFQLDERRRDAVIALGIFLIESDL--	122	QHKDCVVYPYLLRLLKGLPKVYVVEESTARKGR
chimpanze	4	VDF--HGIFQLDERRRDAVIALGIFLIESDL--	64	QHKDCVVYPYLLRLLKGLPKVYVVEESTARKGR
Orang	4	VDF--HGIFQLDERRRDAVIALGIFLIESDL--	64	QHKDCVVYPYLLRLLKGLPKVYVVEESTARKGR
Poule	83	VDF--HGIFQLDERRRDAVIALGIFLIESDL--	143	QHKDNVYPYLLRLLKGLPKVQWIEESSARKGR
Xenopus	4	VDF--HGIFQLDERRRDAVIALGIFLIESDL--	64	QHKESVYPYLLKLLRGLPKVQWIEESSARKGR
poisson_Tetraodon	4	VDF--RGVYQLDERRRDAVIALGIFLVESDLQVLHNISSLLTKHRMKVALVRYSCVTRVTHVLPQHKDVIVPYLLGLLLKGLPKVQWIEESSARKGR	99	LQHKDRILPYLLRLLRGLPKAVWRDEVRCSLAE
drosophile_Drosophila	42	QEN--AAGVFCLODTRAQDAVIALGIYFLESQ--	103	CQHEsqIVPYLLRLAKCLPKAVWIDARSKIE
Zebrafish	4	VDI--RGVFLDERRRDAVIALGIYFLESQ--	64	QHKDAIVPYLLVLLKGLPKVQWIEESSARKGR
guepe	40	QDS--GGVFRLDQRGQDAVIALGIYFLESQ--	100	LQHRDRILPYLLKLLRGLPKAIWLDEAPCLLLG
moustique	40	QEN--AAGVFCLODTRAQDAVIALGIYFLESQ--	101	YQYSKEIIPYLLRLAKALPKAVWIDVKSNNID
foumi_Harpegnathos	40	QDS--AAGVYRLDQRGQDAVIALGIYFLESQ--	101	LQHKDRILPYLLRLLRGLPKAVWRDEVRCSLAE
pou	40	KKYDSEIYHIDQRGQDAVIALGIYFLESQ--	102	FQHSKILPYLLKLLKALHLVKWLDIPLGKDD
tique_kodes	42	QEL--PQGNFRIDQRGQDAVIALGIYFLESQ--	103	FQAKDQIVPYLLRLLRGLPKAVWLDEVHSDQE
crustace_Daphnia	39	ASG--ARGEYKLDHRGQSAVIALGIYFLESRLNF--	107	PNSPFPKHNKILPYFFDLIRNLPLARWTDSQLLFQNR
nematode_Ascaris	39	FDV--PEKLALNHAIROCKLLAAGIYLVYAD--	99	GKHLDELVPFLKIFVSLPQMKWVDDGVANKAD
nematode_Caenorhabditis	39	KEE--QFESQPLSHNVRCSLIGIANFLLNSH--	100	GKFAEELVPMLLRALSALPHMKWIDGGLLNKSD
cerevisiae	76	AIP--	113	TTLDLVETMLKEIIPYFLETP--RQRISDVLSSKF
aspergillus	16	AAP--	61	LVTKIDHASKMVSQSLPYLPESH--SQCFR--PSPF
neurospora	85	AAP--	120	SLTTAQAQKLTRHLVBYLMDSH--SQTF--LSPFF

Consensus

VDF--AHGVFQLDERRRDAVIALGI+LLESDL--VT---S++QHKD+IVPYLLRLLKGLPKVQWIDES+A+KGR

humain	123	-GALPVAESFSFCLVTLSDVAYRDPRLRD-EILEVLLQVLHVLLGMQCALEIQ--	203	DKEYLCKYAIPLCLIGISRAFGRYSN-MEES
chimpanze	65	-GALPVAESFSFCLVTLSDVAYRDPRLRD-EILEVLLQVLHVLLGMQCALEIQ--	145	DKEYLCKYAIPLCLIGISRAFGRYSN-MEES
Orang	65	-GALPVAETFSFCLVTLSDVAYRDPRLRD-EILEALLQVLHVLLGMQCALEIQ--	145	DKEYLCKYAIPLCLIGISRAFGRYSN-MEES
Poule	144	-GLLPVAENFSFCLVTLSDVAYRDASLRE-QLDAVLQVMQLLGMCMQTPQMH--	224	DKEYLCKYAVPCLIGISRAFGRYSN-SEEA
Xenopus	65	-GTLPVAETFSFCLVTLSDVAYRDPRLRD-EILETLMQVLMQCLLGMCMQSPQIQ--	145	DKEYLCKYAVPCLIGISRAFGRYSN-TEES
poisson_Tetraodon	100	-EKLPAENFTFSFCLVTLSDVAYRDPRLRD-EILETLMQVLMQCLLGMCMQSPQIQ--	152	DKEYLCKYAVPCLIGISRAFGRYSN-TEES
drosophile_Drosophila	104	R-VRIPIAEKFSFCLNTLLSDIAAKCPDSRE-EILNQVETLQALANIVKSSRDSSSA--	189	PPPIILCKATVPLLGLARSMGRYAS-NDPP
Zebrafish	65	-EALPIAENFSFCLVTLSDVAYRDPRLRD-EILETLMQVLMQCLLGMCMQSPQIQ--	189	DKEYLCKYAVPCLIGISRAFGRYSN-TEES
guepe	101	-RVPVAERFSFCLNTLLSDIAARCEPARE-EIATQVEILAVLSNLVRGCIQNNGR--GMQAKCTPTLTLCLVPLVGLARSMGRFAC-EGPP	191	PPPIILCKATVPLLGLARSMGRYAS-NDPP
moustique	102	-KIPTAEKFSFCLNTLLSDIASGCEPHRD-EILNQVEVLTVLSTMIKSGKESNT--	184	PPPIILCKATVPLLGLARSMGRYAS-NDPP
foumi_Harpegnathos	102	-RVPVAERFSFCLNTLLSDIAARCEPARE-EIATQVEILAVLSNLVRGCIQNNGR--GMQAKCTPTLTLCLVPLVGLARSMGRFAC-EGPP	188	PPPIILCKATVPLLGLARSMGRYAS-NDPP
pou	103	-RVPVAERFSFCLNTLLSDIAARNDLLRE-RIITAQIKQESLCHLIHKKHNSSTH--	190	TDQSKIYLCRNVPVFFGLARSMGRYST-DPP
tique_kodes	104	-RIPVCEERFAFCLATLLSDIAYLCEELRE-EISAQIDFLAVLVNICKSFADAQTPR--GS--TGKVTLCRSIVPLLIGTARAVGRAKA-SAPP	190	TGKVTLCRSIVPLLIGTARAVGRAKA-SAPP
crustace_Daphnia	108	-CLPTSECFAFSMTLLSDVAVLDEEARA-EILDFQVTLFIEISQLILSKRQDSN--	186	PPNVVLCRIVPPLLFGHARSMGRFSA--P
nematode_Ascaris	100	-RVPVQEQAFACFNTVLSLAFAFYNVRD-QIVSAQLEVLACCTNTIVDLCEMQQP--	185	LLPAKVHLMKMVTFVIGLTRSFARYSGNKETP
nematode_Caenorhabditis	101	-RIPVQEQAFACFNTVLSLAFAFYNVRD-QIVSAQLEVLACCTNTIVDLCEMQQP--	186	LEHNYSHVVKLMQVVLGLTRSGRSTDYDTP
cerevisiae	114	LEQMRNPILLTQLTKFMIAQCEQYPLYE-NIGGIISTYFERYLK--	186	IFTIKQSGLLSLVGFINAFIQFPN--STE
aspergillus	52	YDLKPSPWKLTLYKLTSSLSLGSNFPPLRS-TTLSAINEYLYSCADAIDAITPLQRYDSAGDKFQSVQEHKRLSLAVSLVGLFLEASAOFPS--	143	YDLKPSPWKLTLYKLTSSLSLGSNFPPLRS-TTLSAINEYLYSCADAIDAITPLQRYDSAGDKFQSVQEHKRLSLAVSLVGLFLEASAOFPS--
neurospora	121	REVEPSPTESLSYHVTTALLSLGIGHATHVQVEVSDSIWAYLNACAHATQTLTEANYE--	208	GPAALDDAIRTATITVALLGLDAACAKAD--

Consensus

REGRLPVAERFSFCL+TLLSDVAYRDPRLRD-EILEAQ+EVLAVLL+MCQ+LE+QNSNR--G-Q--LLDKEYLCKYAVPCLIGLARAFGRYSN-TE+P

humain	204	LLSKLFPKIPPHSLRVLEE--	262	LEGVRRRSFNDFRSILPSNLLTVQCEGLTKRKTSSVSSI--
chimpanze	146	LLSKLFPKIPPHSLRVLEE--	204	LEGVRRRSFNDFRSILPSNLLTVQCEGLTKRKTSSVSSI--
Orang	146	LLSKLFPKIPPHSLRVLEE--	204	LEGVRRRSFNDFRSILPSNLLTVQCEGLTKRKTSSVSSI--
Poule	225	LLSKLFPKIPPHSLRVLEE--	283	LEGIRRRSFNDFRSILPSSLLTVQCEGLTKRKTSSVSSI--
Xenopus	146	LLSKLFPKVPSPQSLKAPEE--	204	LEGIRRRSFNDFRSILPSSLLTVQCEGLTKRKTSSVSSI--
poisson_Tetraodon	153		207	KGAEGMAYFHQIPLLTVCQGDTLRRKGPLSSSPSQVEDKNSGNVSKVEVCMDRL
drosophile_Drosophila	190	LLCRIFPPPELLPIQRVSGRDGTASSASGTCGGSFSSSERLAAATANHQRFPPIPRSMSSGLAAQYHQHFDGGRQRLHGLVGFQSN--	274	LLCRIFPPPELLPIQRVSGRDGTASSASGTCGGSFSSSERLAAATANHQRFPPIPRSMSSGLAAQYHQHFDGGRQRLHGLVGFQSN--
Zebrafish	146	LLSKLFPRIIPVPPGPGTG--	205	AEVTRRRSFNDFRSILPSSLLTVQCEGLTKRKTSSVSSI--
guepe	192	LLCRIFPRPEPPIPKCK--TE-SHPL--	258	PDKKQHSFCNFRSIPRSLSGDLNPLRSTGTG--QTSFPQANENNS--
moustique	185	LLCRIFPKDEILVVKSAET--	246	PLCMAD--AKDSITGVQSFRSIPRSMSSGLTIDSTETEKSSS--S--
foumi_Harpegnathos	189	LLCRIFPRRETSPATN--MDHSRSM--	256	PDKKQHSFTNFRPIPRSLSGFNFCIDNTPO--LLSAYDNHDYHF--
pou	191	LQCIIFPKPLPPTQTE--	249	TCL--NLTKKKNFTFRNIIIPHLSANLSDSTFLQG--PNKGHQIE--
tique_kodes	191	LFQCFIPPEKVVLTSEDAPQRSSSSSG--	259	TLTKRRTFNFRPIIPQSLSELSQLAHGRDK--SYEDDDEV--
crustace_Daphnia	189	LFKLFPKATPPVLRTEP--	239	SSTKNRSFNFRSIPQSLSTTIFANIDPFHS--
nematode_Ascaris	186	LISITPIIPFLSER--	210	DESCLDEEQ--
nematode_Caenorhabditis	187	LISLIFPLPMDETKKLGKNG--	216	AEDCLHRIS--
cerevisiae	187	LTKFTWKKLAKLVLRG--	204	S--
aspergillus	144	LWSASEKQIV--	174	HVRSMLSEGFMAIETASS--
neurospora	209	FWKSGGRLSLVE--	239	RLRNMLSPFFLVAVETAFS--

Consensus

LLSKLFPK+EP+SLRVLEE--SSSSS+LCL--SSLEG+RRRSFNDFRSILPSSLLTVQCE+TLTKRKTSSVSSI+S+-D-NS--









**ANNEXE 4 : EXPRESSION EN BACTERIE DES CONSTRUCTIONS C-TERMINALES ET SOLUBILITE**

Construction	Vecteur	Bactérie			Lyse	Expression	Solubilité
		souche	T°C	IPTG			
<b>1472-2102</b>	Pgex6p1 Tag GST	BL21(DE3)	18°C	1mM	0.5% Triton X-100	++	peu
		BL21(DE3)Rosetta				+++	
		BL21(DE3)Star				0	
		BL21(DE3)C41	12°C	0.3mM	0.5% triton	++	peu
					43.5mM LDAO	++	+++
			18°C	1mM	0.5% triton	+	peu
<b>1534-2102</b>	Pgex6p1 Tag GST	BL21(DE3)	18°C	1mM	0.5% Triton X-100	+++	25%
		BL21(DE3)Rosetta				++	75%
		BL21(DE3)Star				++	50%
		Pgro7	12°C	0.3		Ø	/
		Arctic express	12°C	1mM	Sans détergent	+	+ +chaperonne
		BL21(DE3) C41	18	1mM	0.5% triton	++	25%
			16		1	++	25%
					0.5	++	peu
			12	0.3	43.6 LDAO	++	++
				0.1		Ø	Ø
	6HIS Nter	BL21(DE3) C41	12°C	0.3	1% triton	Ø	/
					Sans triton	Ø	/
<b>1534-2102 dc17</b>	Pgex6p1 Tag GST	BL21(DE3) C41	16°C	1mM	1% triton	+	peu
<b>1508-2102</b>	Pgex6p1 Tag GST	BL21(DE3)	18°C	1mM	0.5% Triton X-100	+++	peu
		BL21(DE3)Rosetta				++	50%
		BL21(DE3)Star				+++	peu
		BL21(DE3)C41	16	1		+++	peu
					1%	+++	20%
					Sans triton	+++	peu
			12	0.3	0.5% Triton	+++	25%
					43.6 mM LDAO	+++	+++
					0.5%	+++	peu
<b>1508-2102dc17</b>	Pgex6p1 Tag GST	BL21(DE3) C41	16	1	1% triton	+	25%
	6his	BL21(DE3) C41	12	0.3	1% triton	Ø	/
<b>1293-2102</b>	Pgex6p1 Tag GST	BL21(DE3) C41	18	1	Sans detergent	+	peu
	6his	BL21(DE3)Rosetta	12	0.3	43.6mM LDAO	+	+
		pGro7				Ø	Ø
<b>1135-2102</b>	Pge6p1	BL21(DE3) C41	12	0.3	Sans detergent	+	peu
		Pgro7	12	0.3	43.6 mM LDAO	++	peu

**ANNEXE 5 : PURIFICATION DES CONSTRUCTIONS C-TERMINALES**

<b>Construction</b>	<b>vecteur</b>	<b>détergent Cofacteur</b>	<b>Chromatographie d'affinité</b>	<b>Coupure étiquette</b>	<b>Gel filtration</b>	<b>Concentration</b>
<b>1472-2102</b>	GST	Ø detergent 1mM MgCl <sub>2</sub>	✓	50%	V <sub>m</sub> sur S200	Perte 50% 1.44 mg/mL
		0.1% triton X100 1mM MgCl <sub>2</sub>	✓	50%	S200 V <sub>e</sub> = 10.94	Ne se concentre pas
		0.25 mM triton	25% éluee	/	/	/
		0.4% lauryl sarcosyl 1mM MgCl <sub>2</sub>	Ø élution	/	/	/
		2 mM LDAO	✓	Ø	V <sub>e</sub> =9.96mL superose 6 10/300	/
<b>1534-2102</b>	GST	Ø detergent	Ø élution	/	/	/
		0.25 mM Triton	25% éluee	/	/	/
		5mM LDAO	✓	/	/	/
<b>1534- 2102dc17</b>	GST	Ø detergent	Ø élution	10%	V <sub>m</sub>	Perte 50%
		1% triton	Ø élution			
<b>1508-2102</b>	gst	Ø detergent	Peu éluee	75%	V <sub>m</sub>	/
		0.25 mM triton	✓	/	V <sub>e</sub> =12.79 S200 avec dna K	/
		0.25 mM triton 0.1mM ATP	✓	50%	V <sub>e</sub> =12.77 S200 avec dna K	
		100 mM NaNO <sub>3</sub>	✓	Ø	V <sub>m</sub>	
		5mM LDAO	✓	/	/	/
<b>1508- 2102dc17</b>	gst	Ø detergent	Ø élution	/	/	/
<b>1135-2102</b>	GST	2mM LDAO	✓	/	/	/
		Ø detergent	✓	✓	/	/
<b>1293-2102</b>	6his nter	Ø detergent	✓	/	/	/
	gst	Ø detergent	✓	✓	/	/

**ANNEXE 6 : CONDITIONS D'EXPRESSION TESTEES DES CONSTRUCTIONS N-TERMINALES EN BACTERIE ET SOLUBILITE**

Construction	Vecteur	Bactérie			Lyse	Expression	Solubilité	
		souche	T°C	IPTG				
611-1088	pGex6p1	BL21DE3C41	16°C	1mM	1% triton	+++	++	
					sans détergent	+++	++	
				0.3mM	1% triton	++	50%	
			18°C	0.3mM	1% triton	++	50%	
			12°C	0.3mM	1%triton	++++	50%	
	Pet28	BL21DE3C41	16°C	1mM		Ø		
611-1118	Pgex6p1	BL21DE3C41	16°C	1mM	1% triton	+++	++	
			12°C	0.3mM	1% triton	++++		
	Pet28	BL21DE3C41	16°C	1mM	/	Ø	/	
553-1118	Pgex6p1	Rosetta	18°C	0.1	/	Ø	/	
		C41	18°C	1mM	Sans détergent	+++	40%	
					0.5% triton	+++	50%	
	Pet 28	C41	18°C	0.1	Sans détergent	Ø		
553-1275	pgex	C41	18°C	1	Sans détergent	+++	faible	
276-1275	Pgex6p1	Rosetta	18°C	1	sans détergent	+++	faible	
				0.1		+++	faible	
				1		+++	faible	
				0.1		Ø	/	
		Pet28		Rosetta		1	++	faible
		0.1				++	faible	
		1				++	faible	
		0.1				++	faible	
	C41							
276-1118	Pgex6p1	C41	18°C	1	Sans détergent	+++	faible	
					0.5% triton	+++	+	
26-1118	Pgex6p1	C41	18°C	1	Sans détergent	+++	25%	
					0.5% triton	+++	++	
				23°C	1	Sans détergent	+++	++
26-1275		Pgex6p1	C41	18°C	1	Sans détergent	+++	faible
1-1275	pGex6p1	Rosetta	18°C	0.1	Sans détergent	+/-	faible	
				1		+	faible	
		C41		0.1		Ø	/	
				1		+	faible	

**ANNEXE 7 : TESTS DE PURIFICATION DES CONSTRUCTIONS N-TERMINALES**

<b>Construction</b>	<b>vecteur</b>	<b>détergent Cofacteur</b>	<b>Chromatographie d'affinité</b>	<b>Coupure étiquette</b>	<b>Gel filtration</b>	<b>concentration</b>
<b>611-1088</b>	Pgex6p 1	0.25 mM triton	✓	✓	S200 10/300GL Ve 13.18mL MW = 121kDa Rs=4,07nm	
		Sans détergent Expression 18°C	Non retenu	Pas de coupure		
		Sans détergent Expression 12°C	✓	Coupure mais agrège		
<b>611-1118</b>	Pgex6p 1	0.25 mM triton	✓			
		Sans détergent	✓	Coupure mais agrège		
<b>553-1118</b>	Pgex6p 1	Sans détergent	✓	✓	superose 6 100 mM NaCl Ve = 11.72 mL MW = Rs=	
<b>276-1118</b>	Pgex6p 1	Sans détergent	Elution partielle	✓	Pas testé	
<b>25-1118</b>	Pgex6p 1	Sans détergent	✓	✓	Sup6 Ve=15.73mL MW Rs	✓

**ANNEXE 8 : COMPOSITION DES DIFFERENTS MILIEUX DE CULTURE UTILISES POUR LA MISE AU POINT DE  
L'EXPRESSION DE 26-1118**

	<b>Lysogeni broth <i>industriel</i></b>	<b>YT <i>industriel</i></b>	<b>Milieu Auto-inductible</b>
	10g/L biotrypcase 5g/L extrait de levure	16g/L tryptone 10g/L d'extrait de levur	bacto-tryptone extrait de levure
<b>Sels et miénraux</b>	5g/L NaCl	5 g/L NaCl	MgSO4 (NH4)2SO4 KH2PO4 Na2HPO4
<b>Vitamine</b>			
<b>Substrats carbonés</b>			10g/L glycérol 1g/L glucose 4g/L gamma lactose

## ANNEXE 9 : MISE AU POINT DE LA PURIFICATION DE 26-1118

*La concentration en NaCl et en glycérol correspond à celle tiliser lors de la purification après lea fiation à la résine glutathione Sépharose 4B.*

[NaCl]	glycérol	Chromatographie d'affinité	Coupure étiquette	Gel filtration	Stabilité
100 mM	5%	Peu de protéine	✓	Pas ressortie	/
300mM	5%	Peu de protéine	✓	✓	/
500mM	10%	✓	✓	✓	Perte haute [c]
1M	10%	✓	✓	✓	Peu stable dans le temps



**ANNEXE 10 : TESTS INTERACTION ENTRE DI DE PI4KA ET D1-NS5A**

		<b>Incubation de NS5A avec DI</b>	<b>Tampon</b>	<b>Ratio NS5A : DI</b>	<b>interaction</b>
Pull down	GST	30 minutes avant résine	15 mM Tris pH 7,4 25mM Phosphate pH 7,5 300 mM NaCl 1mM LDAO 5% glycérol	/	Une petite partie de GST-DI dans le non retenu
		Sur la nuit	12,5 mM Hepes 15 mM Tris pH 7,4 200 mM NaCl 10% glycérol	Excès NS5A	
		1 heure avec résine 45 minutes avant résine 4 heures avec résine	12,5 mM Hepes 15 mM Tris pH 7,4 200 mM NaCl 10% glycérol 2 mM LDAO	1 : 2	
	Nickel	1 heure avec résine 45 minutes avant résine 2 heures avec résine	12,5 mM Hepes 15 mM Tris pH 7,4 200 mM NaCl 10% glycérol 2 mM LDAO	Excès NS5A 2 : 1	Interaction GST-DI avec la résine  Une partie de d1-NS5A dans le non retenu
Crosslink		1 heure	30 mM Hepes pH 7,4 200 mM NaCl 15% glycérol 1 mM LDAO	2 : 1	Pas d'interaction
		1 heure 3 heures	30 mM Hepes pH 7,4 200 mM NaCl 15% glycérol 2 mM LDAO		
Gel filtration		2 heures	30mM Tris pH 7,4 300 mM NaCl 2mM LDAO 10% glycérol 1 mM DTT	1 : 1	Pas d'interaction
Co-expression	/	/	/	Pas d'expression de DI sans étiquette	/
Co-purification		/	25 mM Hepes pH 7,4 100 mM NaCl 20% glycérol	/	Pas d'interaction

**ANNEXE 11. PROTOCOLE D'EXPRESSION ET DE PURIFICATION DE LA PI4KA HUMAINE ENTIERE MIS AU POINT  
PAR L'EQUIPE DE VOLKER LOHMANN**

- infection of Sf9 cells with recombinant Baculovirus expressing PI4KIIIa and GFP
- expansion of cell culture system to needed volume
- harvest cells by scrapping them off in cold PBS (3ml per 15cm dish)
- centrifuge supernatant and add cell pellet to scrapped cell pellet (freeze pellet at -70°C or lyse directly)
- use cleared supernatant for further infections

Purification performed on ice/coldroom!!!

- lyse cell pellet in Lysis Buffer (2) 10 min on ice: 1ml per 10cm dish ; 2,5ml per 15cm dish
- 4x freeze and thaw in liquid N2 (take T = 50µl)
- add same volume of Dilution Buffer (5) take aliquot (150µl → 5min/max. rpm) for control gel (→ S1 and pellet)
- centrifuge at 4°C 14,000rpm for 30min
- incubate (shake) supernatant (S1) with 1ml Nickel Beads Slurry (10x15cm dishes), 200µl for 1x15cm dish, 100µl for 1x10cm dish for 2hours or over night at 4°C
- centrifuge at 1,000 g for 2-3 min (4°C) and take aliquot for control gel (FT1; 50µl)
- freeze first flow through (FT1) (optional: with 10% glycerol)
- wash 7x with Wash Buffer (3) and 1x with Basis-Buffer for purification (1)
- transfer beads to Elution Column and wash again with 10ml Basis-Buffer for purification (1)
- elute with Elution buffer (4)
  - E1-E3: half volume of used Beads Slurry; E4: ¾ volume of used Beads Slurry; take aliquots (10µl for control gel)

<u>Basis-Buffer for purification (1)</u>	<u>Lysis Buffer (2)</u>	<u>Wash Buffer (3)</u>	<u>Elution Buffer (4)</u>	<u>Dilution Buffer (5)</u>
50mM KH <sub>2</sub> PO <sub>4</sub> pH 7.4 300mM NaCl	add 1% Triton X 100 PMSF(Stock 200mM) 1/200 Aprotenin(1mg/ml) 1/1000 Leupeptin 1/1000	50mM KH <sub>2</sub> PO <sub>4</sub> pH 7.4 1M NaCl 0,5% Triton X 100 25mM Imidazol 14µl beta-ME per 10ml	50mM KH <sub>2</sub> PO <sub>4</sub> pH 7.4 150mM NaCl 50% Glycerol (99,99%) 0,5% Triton X 100 300mM Imidazol 14µl beta-ME per 10ml	10mM KH <sub>2</sub> PO <sub>4</sub> pH 7

## **Hepatitis C Virus promotes the accumulation of lipid droplets through a Septin 9-PtdIns5P-ERK cascade**

Akil Abdallah<sup>1,2,7\*</sup>, Juan Peng<sup>1,2,\*</sup>, Mohyeddine Omrane<sup>1,2\*</sup>, Claire Gondeau<sup>3,4,5</sup>, Mickaël Marin<sup>1,2</sup>, Hélène Tronchère, Cyntia Taveneau<sup>6</sup>, Célia-Caillet-Saguy<sup>6,9</sup>, Sokhavuth Sar<sup>1,2</sup>, Philippe Briolotti<sup>3,4,5</sup>, Soumaya Benjelloun<sup>7</sup>, Abdelaziz Benjouad<sup>8</sup>, Patrick Maurel<sup>3,4,5</sup>, Valérie Thiers, Stéphane Bressanelli<sup>6</sup>, Didier Samuel<sup>1,2</sup>, Christian Bréchet<sup>1,2,9</sup> and Ama Gassama-Diagne<sup>1,2</sup>

### **SUMMARY**

Liver steatosis characterized by the accumulation of lipid droplets (LDs) is a frequent feature of chronic hepatitis C virus (HCV) infection and an important risk factor for the development of liver cirrhosis and cancer. The mechanisms controlling LDs biogenesis, their cellular dynamic and function remain unsolved issues. Here, we demonstrate that septin 9 regulates HCV replication and the clustering of LDs. Septin 9 binds to monophosphorylated phosphoinositides (PIs), mostly PtdIns5P, and both form ring like structures surrounding LDs. PtdIns5P regulates the clustering of LDs, assembly of septin 9 into rings and filaments through ERK 1/2 phosphorylation. This regulatory role of PtdIns5P is confirmed using a septin 9 mutant incapable of binding to PIs. Furthermore we show that PtdIns5P, septin 9 and ERK2 are required for oleate- induced LDs accumulation. Overall, our data offer a novel paradigm for the formation of LDs through the unexpected implication of PtdIns5P signaling pathway.



# Bibliographie

---

1. OMS | Hépatite C. WHO at <<http://www.who.int/mediacentre/factsheets/fs164/fr/>>
2. Ansaldi, F. (2014). Hepatitis C virus in the new era: Perspectives in epidemiology, prevention, diagnostics and predictors of response to therapy. *World Journal of Gastroenterology* 20, 9633.
3. Averhoff, F.M., Glass, N., and Holtzman, D. (2012). Global burden of hepatitis C: considerations for healthcare providers in the United States. *Clin. Infect. Dis.* 55 Suppl 1, S10–15.
4. Moradpour, D., Penin, F., and Rice, C.M. (2007). Replication of hepatitis C virus. *Nature Reviews Microbiology* 5, 453–463.
5. Lee, M.-H., Yang, H.-I., Yuan, Y., L'Italien, G., and Chen, C.-J. (2014). Epidemiology and natural history of hepatitis C virus infection. *World J. Gastroenterol.* 20, 9270–9280.
6. Maasoumy, B., and Wedemeyer, H. (2012). Natural history of acute and chronic hepatitis C. *Best Pract Res Clin Gastroenterol* 26, 401–412.
7. Hoofnagle, J.H. (1999). Management of hepatitis C: current and future perspectives. *J. Hepatol.* 31 Suppl 1, 264–268.
8. Sulkowski, M.S. (2014). Interferon-containing and interferon-free HCV therapy for HIV-infected patients. *Semin. Liver Dis.* 34, 72–78.
9. ANRS | <http://www.anrs.fr/Hepatitis-virales-B-et-C/Enjeux-de-la-recherche>
10. Pol, S., and Corouge, M. (2014). Treatment of hepatitis C: Perspectives. *Med Mal Infect.*
11. Messina, J.P., Humphreys, I., Flaxman, A., Brown, A., Cooke, G.S., Pybus, O.G., and Barnes, E. (2014). Global distribution and prevalence of hepatitis C virus genotypes. *Hepatology*.
12. Rice, C.M., and Saeed, M. (2014). Hepatitis C: Treatment triumphs. *Nature* 510, 43–44.
13. ANRS | <http://www.anrs.fr/Hepatitis-virales-B-et-C/Clinique/Actualites/Eradiquer-l-hepatite-C-mais-a-quel-prix>
14. International Committee of Taxonomy of Viruses
15. Hnatyszyn, H.J. (2005). Chronic hepatitis C and genotyping: the clinical significance of determining HCV genotypes. *Antivir. Ther. (Lond.)* 10, 1–11.
16. Simmonds, P. (2013). The origin of hepatitis C virus. *Curr. Top. Microbiol. Immunol.* 369, 1–15.
17. Smith, D.B., Bukh, J., Kuiken, C., Muerhoff, A.S., Rice, C.M., Stapleton, J.T., and Simmonds, P. (2014). Expanded classification of hepatitis C virus into 7 genotypes and 67 subtypes: Updated criteria and genotype assignment web resource: *Hepatology*, Vol. 00, No. X, 2013 Smith et al. *Hepatology* 59, 318–327.
18. Bartenschlager, R., Lohmann, V., and Penin, F. (2013). The molecular and structural basis of advanced antiviral therapy for hepatitis C virus infection. *Nat. Rev. Microbiol.* 11, 482–496.
19. Duffy, S., Shackelton, L.A., and Holmes, E.C. (2008). Rates of evolutionary change in viruses: patterns and determinants. *Nat. Rev. Genet.* 9, 267–276.
20. Yu, X., Qiao, M., Atanasov, I., Hu, Z., Kato, T., Liang, T.J., and Zhou, Z.H. (2007). Cryo-electron microscopy and three-dimensional reconstructions of hepatitis C virus particles. *Virology* 367, 126–134.

21. Lindenbach, B.D., and Rice, C.M. (2013). The ins and outs of hepatitis C virus entry and assembly. *Nature Reviews Microbiology* 11, 688–700.
22. Pécheur, E.-I. (2012). Lipoprotein Receptors and Lipid Enzymes in Hepatitis C Virus Entry and Early Steps of Infection. *Scientifica* 2012, 1–11.
23. Kim, C.W., and Chang, K.-M. (2013). Hepatitis C virus: virology and life cycle. *Clin Mol Hepatol* 19, 17–25.
24. McMullan, L.K., Grakoui, A., Evans, M.J., Mihalik, K., Puig, M., Branch, A.D., Feinstone, S.M., and Rice, C.M. (2007). Evidence for a functional RNA element in the hepatitis C virus core gene. *Proc. Natl. Acad. Sci. U.S.A.* 104, 2879–2884.
25. Altan-Bonnet, N., and Balla, T. (2012). Phosphatidylinositol 4-kinases: hostages harnessed to build panviral replication platforms. *Trends Biochem. Sci.* 37, 293–302.
26. Chatel-Chaix, L., and Bartenschlager, R. (2014). Dengue virus- and hepatitis C virus-induced replication and assembly compartments: the enemy inside--caught in the web. *J. Virol.* 88, 5907–5911.
27. Romero-Brey, I., Merz, A., Chiramel, A., Lee, J.-Y., Chlanda, P., Haselman, U., Santarella-Mellwig, R., Habermann, A., Hoppe, S., Kallis, S., et al. (2012). Three-Dimensional Architecture and Biogenesis of Membrane Structures Associated with Hepatitis C Virus Replication. *PLoS Pathog* 8, e1003056.
28. Reiss, S., Harak, C., Romero-Brey, I., Radujkovic, D., Klein, R., Ruggieri, A., Rebhan, I., Bartenschlager, R., and Lohmann, V. (2013). The lipid kinase phosphatidylinositol-4 kinase III alpha regulates the phosphorylation status of hepatitis C virus NS5A. *PLoS Pathog.* 9, e1003359.
29. Farci, P. (2002). Choo QL, Kuo G, Weiner AJ, Overby LR, Bradley DW, Houghton M. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome [Science 1989;244:359-362]. *J. Hepatol.* 36, 582–585.
30. S., Rebhan, I., Backes, P., Romero-Brey, I., Erfle, H., Matula, P., Kaderali, L., Poenisch, M., Blankenburg, H., Hiet, M.-S., et al. (2011). Recruitment and activation of a lipid kinase by hepatitis C virus NS5A is essential for integrity of the membranous replication compartment. *Cell Host Microbe* 9, 32–45.
31. Tai, A.W., and Salloum, S. (2011). The role of the phosphatidylinositol 4-kinase PI4KA in hepatitis C virus-induced host membrane rearrangement. *PLoS ONE* 6, e26300.
32. Balla, T. (2013). Phosphoinositides: tiny lipids with giant impact on cell regulation. *Physiol. Rev.* 93, 1019–1137.
33. Falkenburger, B.H., Jensen, J.B., Dickson, E.J., Suh, B.-C., and Hille, B. (2010). Phosphoinositides: lipid regulators of membrane proteins. *J Physiol* 588, 3179–3185.
34. Stahelin, R.V., Scott, J.L., and Frick, C.T. (2014). Cellular and molecular interactions of phosphoinositides and peripheral proteins. *Chem. Phys. Lipids* 182, 3–18.
35. Di Paolo, G., and De Camilli, P. (2006). Phosphoinositides in cell regulation and membrane dynamics. *Nature* 443, 651–657.
36. D'Angelo, G., Vicinanza, M., Di Campli, A., and De Matteis, M.A. (2008). The multiple roles of PtdIns(4)P -- not just the precursor of PtdIns(4,5)P2. *J. Cell. Sci.* 121, 1955–1963.
37. Billcliff, P.G., and Lowe, M. (2014). Inositol lipid phosphatases in membrane trafficking and human disease. *Biochem. J.* 461, 159–175.
38. Cell Signalling Biology - Michael J. Berridge - [www.cellsignallingbiology.org](http://www.cellsignallingbiology.org) – 2014
39. Graham, T.R., and Burd, C.G. (2011). Coordination of Golgi functions by phosphatidylinositol 4-kinases. *Trends Cell Biol.* 21, 113–121.



40. Hammond, G.R.V., Fischer, M.J., Anderson, K.E., Holdich, J., Koteci, A., Balla, T., and Irvine, R.F. (2012). PI4P and PI(4,5)P<sub>2</sub> are essential but independent lipid determinants of membrane identity. *Science* 337, 727–730.
41. Lukacs, V., Thyagarajan, B., Varnai, P., Balla, A., Balla, T., and Rohacs, T. (2007). Dual regulation of TRPV1 by phosphoinositides. *J. Neurosci.* 27, 7070–7080.
42. Hanada, K., Kumagai, K., Yasuda, S., Miura, Y., Kawano, M., Fukasawa, M., and Nishijima, M. (2003). Molecular machinery for non-vesicular trafficking of ceramide. *Nature* 426, 803–809.
43. Balla, T., Varnai, P., Tian, Y., and Smith, R.D. (1998). Signaling events activated by angiotensin II receptors: what goes before and after the calcium signals. *Endocr. Res.* 24, 335–344.
44. Szentpetery, Z., Szakacs, G., Bojjireddy, N., Tai, A.W., and Balla, T. (2011). Genetic and functional studies of phosphatidylinositol 4-kinase type III $\alpha$ . *Biochim. Biophys. Acta* 1811, 476–483.
45. Kohout, S.C., Bell, S.C., Liu, L., Xu, Q., Minor, D.L., and Isacoff, E.Y. (2010). Electrochemical coupling in the voltage-dependent phosphatase Ci-VSP. *Nat. Chem. Biol.* 6, 369–375.
46. Tabuchi, M., Audhya, A., Parsons, A.B., Boone, C., and Emr, S.D. (2006). The phosphatidylinositol 4,5-bisphosphate and TORC2 binding proteins Slm1 and Slm2 function in sphingolipid regulation. *Mol. Cell. Biol.* 26, 5861–5875.
47. Audhya, A., Foti, M., and Emr, S.D. (2000). Distinct roles for the yeast phosphatidylinositol 4-kinases, Stt4p and Pik1p, in secretion, cell growth, and organelle membrane dynamics. *Mol. Biol. Cell* 11, 2673–2689.
48. Trotter, P.J., Wu, W.I., Pedretti, J., Yates, R., and Voelker, D.R. (1998). A genetic screen for aminophospholipid transport mutants identifies the phosphatidylinositol 4-kinase, STT4p, as an essential component in phosphatidylserine metabolism. *J. Biol. Chem.* 273, 13189–13196.
49. Wong, K., Meyers ddR, null, and Cantley, L.C. (1997). Subcellular locations of phosphatidylinositol 4-kinase isoforms. *J. Biol. Chem.* 272, 13236–13241.
50. Nakatsu, F., Baskin, J.M., Chung, J., Tanner, L.B., Shui, G., Lee, S.Y., Pirruccello, M., Hao, M., Ingolia, N.T., Wenk, M.R., et al. (2012). PtdIns4P synthesis by PI4KIII $\alpha$  at the plasma membrane and its impact on plasma membrane identity. *J. Cell Biol.* 199, 1003–1016.
51. Nakagawa, T., Goto, K., and Kondo, H. (1996). Cloning, expression, and localization of 230-kDa phosphatidylinositol 4-kinase. *J. Biol. Chem.* 271, 12088–12094.
52. Kakuk, A., Friedländer, E., Vereb, G., Kása, A., Balla, A., Balla, T., Heilmeyer, L.M.G., Gergely, P., and Vereb, G. (2006). Nucleolar localization of phosphatidylinositol 4-kinase PI4K230 in various mammalian cells. *Cytometry A* 69, 1174–1183.
53. Dumaresq-Doiron, K., Savard, M.-F., Akam, S., Costantino, S., and Lefrancois, S. (2010). The phosphatidylinositol 4-kinase PI4KIII $\alpha$  is required for the recruitment of GBF1 to Golgi membranes. *J. Cell. Sci.* 123, 2273–2280.
54. Wu, X., Chi, R.J., Baskin, J.M., Lucast, L., Burd, C.G., De Camilli, P., and Reinisch, K.M. (2014). Structural insights into assembly and regulation of the plasma membrane phosphatidylinositol 4-kinase complex. *Dev. Cell* 28, 19–29.
55. D.A. Fruman, R.E. Meyers, L.C. Cantley, Phosphoinositide kinases, *Annu. Rev. Biochem.* 67 (1998) 481–507. doi:10.1146/annurev.biochem.67.1.481.
56. J.U. Flanagan, P.R. Shepherd, Structure, function and inhibition of the phosphoinositide 3-kinase p110 $\alpha$  enzyme, *Biochem. Soc. Trans.* 42 (2014) 120–124. doi:10.1042/BST20130255.
57. E. Burke, A.J. Inglis, O. Perisic, G.R. Masson, S.H. McLaughlin, F. Rutaganira, et al., Structures of PI4KIII $\beta$  complexes show simultaneous recruitment of Rab11 and its effectors, *Science*. 344 (2014) 1035–1038. doi:10.1126/science.1253397.

58. Miller, S., Tavshanjian, B., Oleksy, A., Perisic, O., Houseman, B.T., Shokat, K.M., and Williams, R.L. (2010). Shaping Development of Autophagy Inhibitors with the Structure of the Lipid Kinase Vps34. *Science* 327, 1638–1642.
59. Balla, A., Tuymetova, G., Toth, B., Szentpetery, Z., Zhao, X., Knight, Z.A., Shokat, K., Steinbach, P.J., and Balla, T. (2008). Design of drug-resistant alleles of type-III phosphatidylinositol 4-kinases using mutagenesis and molecular modeling. *Biochemistry* 47, 1599–1607.
60. Heilmeyer, L.M.G., Vereb, G., Vereb, G., Kakuk, A., and Szivák, I. (2003). Mammalian phosphatidylinositol 4-kinases. *IUBMB Life* 55, 59–65.
61. Alaimo, P.J., Knight, Z.A., and Shokat, K.M. (2005). Targeting the gatekeeper residue in phosphoinositide 3-kinases. *Bioorg. Med. Chem.* 13, 2825–2836.
62. Gehrmann, T., Gülkan, H., Suer, S., Herberg, F.W., Balla, A., Vereb, G., Mayr, G.W., and Heilmeyer, L.M. (1999). Functional expression and characterisation of a new human phosphatidylinositol 4-kinase PI4K230. *Biochim. Biophys. Acta* 1437, 341–356.
63. Nakagawa, T., Goto, K., and Kondo, H. (1996). Cloning, expression, and localization of 230-kDa phosphatidylinositol 4-kinase. *J. Biol. Chem.* 271, 12088–12094.
64. Chung, J., Nakatsu, F., Baskin, J.M., and De Camilli, P. (2015). Plasticity of PI4KIII $\alpha$  interactions at the plasma membrane. *EMBO Rep.*
65. Tellinghuisen, T.L., Marcotrigiano, J., Gorbalenya, A.E., and Rice, C.M. (2004). The NS5A protein of hepatitis C virus is a zinc metalloprotein. *J. Biol. Chem.* 279, 48576–48587.
66. Tellinghuisen, T.L., Marcotrigiano, J., and Rice, C.M. (2005). Structure of the zinc-binding domain of an essential component of the hepatitis C virus replicase. *Nature* 435, 374–379.
67. Love, R.A., Brodsky, O., Hickey, M.J., Wells, P.A., and Cronin, C.N. (2009). Crystal structure of a novel dimeric form of NS5A domain I protein from hepatitis C virus. *J. Virol.* 83, 4395–4403.
68. Appel, N., Schaller, T., Penin, F., and Bartenschlager, R. (2006). From structure to function: new insights into hepatitis C virus RNA replication. *J. Biol. Chem.* 281, 9833–9836.
69. Liang, Y., Ye, H., Kang, C.B., and Yoon, H.S. (2007). Domain 2 of nonstructural protein 5A (NS5A) of hepatitis C virus is natively unfolded. *Biochemistry* 46, 11550–11558.
70. Verdegem, D., Badillo, A., Wieruszkeski, J.-M., Landrieu, I., Leroy, A., Bartenschlager, R., Penin, F., Lippens, G., and Hanouille, X. (2011). Domain 3 of NS5A protein from the hepatitis C virus has intrinsic alpha-helical propensity and is a substrate of cyclophilin A. *J. Biol. Chem.* 286, 20441–20454.
71. De Chasse, B., Navratil, V., Tafforeau, L., Hiet, M.S., Aublin-Gex, A., Agaugué, S., Meiffren, G., Pradezynski, F., Faria, B.F., Chantier, T., et al. (2008). Hepatitis C virus infection protein network. *Mol. Syst. Biol.* 4, 230.
72. Kaneko, T., Tanji, Y., Satoh, S., Hijikata, M., Asabe, S., Kimura, K., and Shimotohno, K. (1994). Production of two phosphoproteins from the NS5A region of the hepatitis C viral genome. *Biochem. Biophys. Res. Commun.* 205, 320–326.
73. Kakuk, A., Friedländer, E., Vereb, G., Kása, A., Balla, A., Balla, T., Heilmeyer, L.M.G., Gergely, P., and Vereb, G. (2006). Nucleolar localization of phosphatidylinositol 4-kinase PI4K230 in various mammalian cells. *Cytometry A* 69, 1174–1183.
74. Graaf, P. de, Klapisz, E.E., Schulz, T.K.F., Cremers, A.F.M., Verkleij, A.J., and Henegouwen, P.M.P. van B. en (2002). Nuclear localization of phosphatidylinositol 4-kinase  $\beta$ . *J Cell Sci* 115, 1769–1775.
75. Ahn, J., Chung, K.-S., Kim, D.-U., Won, M., Kim, L., Kim, K.-S., Nam, M., Choi, S.-J., Kim, H.-C., Yoon, M., et al. (2004). Systematic identification of hepatocellular proteins interacting with NS5A of the hepatitis C virus. *J. Biochem. Mol. Biol.* 37, 741–748.

76. Balla, A., and Balla, T. (2006). Phosphatidylinositol 4-kinases: old enzymes with emerging functions. *Trends Cell Biol.* 16, 351–361.
77. Lempiäinen, H., and Halazonetis, T.D. (2009). Emerging common themes in regulation of PIKKs and PI3Ks. *EMBO J.* 28, 3067–3073.
78. Lim, Y.-S., and Hwang, S.B. (2011). Hepatitis C virus NS5A protein interacts with phosphatidylinositol 4-kinase type III $\alpha$  and regulates viral propagation. *J. Biol. Chem.* 286, 11290–11298.
79. Baretic, Domagoj, and Roger L. Williams. “PIKKs--the Solenoid Nest Where Partners and Kinases Meet.” *Current Opinion in Structural Biology* 29 (December 2014): 134–42. doi:10.1016/j.sbi.2014.11.003.
80. Sommer, Lisa A. M., Martin Schaad, and Sonja A. Dames. “NMR- and Circular Dichroism-Monitored Lipid Binding Studies Suggest a General Role for the FATC Domain as Membrane Anchor of Phosphatidylinositol 3-Kinase-Related Kinases (PIKK).” *The Journal of Biological Chemistry* 288, no. 27 (July 5, 2013): 20046–63. doi:10.1074/jbc.M113.467233.
81. Ascher, David B., Jerome Wielens, Tracy L. Nero, Larissa Doughty, Craig J. Morton, and Michael W. Parker. “Potent Hepatitis C Inhibitors Bind Directly to NS5A and Reduce Its Affinity for RNA.” *Scientific Reports* 4 (2014): 4765. doi:10.1038/srep04765.
82. Chukkapalli, Vineela, Kristi L. Berger, Sean M. Kelly, Meryl Thomas, Alexander Deiters, and Glenn Randall. “Daclatasvir Inhibits Hepatitis C Virus NS5A Motility and Hyper-Accumulation of Phosphoinositides.” *Virology* 476C (December 26, 2014): 168–79. doi:10.1016/j.virol.2014.12.018.
83. Bressanelli, S., L. Tomei, A. Roussel, I. Incitti, R. L. Vitale, M. Mathieu, R. De Francesco, and F. A. Rey. “Crystal Structure of the RNA-Dependent RNA Polymerase of Hepatitis C Virus.” *Proceedings of the National Academy of Sciences of the United States of America* 96, no. 23 (November 9, 1999): 13034–39.
84. Harrus, Déborah, Neveen Ahmed-El-Sayed, Philip C. Simister, Steve Miller, Martine Triconnet, Curt H. Hagedorn, Kathleen Mahias, Félix A. Rey, Thérèse Astier-Gin, and Stéphane Bressanelli. “Further Insights into the Roles of GTP and the C Terminus of the Hepatitis C Virus Polymerase in the Initiation of RNA Synthesis.” *The Journal of Biological Chemistry* 285, no. 43 (October 22, 2010): 32906–18. doi:10.1074/jbc.M110.151316.
85. Simister, Philip, Melanie Schmitt, Matthis Geitmann, Oliver Wicht, U. Helena Danielson, Rahel Klein, Stéphane Bressanelli, and Volker Lohmann. “Structural and Functional Analysis of Hepatitis C Virus Strain JFH1 Polymerase.” *Journal of Virology* 83, no. 22 (November 2009): 11926–39. doi:10.1128/JVI.01008-09.
86. Mosley, Ralph T., Thomas E. Edwards, Eisuke Murakami, Angela M. Lam, Rena L. Grice, Jinfa Du, Michael J. Sofia, Philip A. Furman, and Michael J. Otto. “Structure of Hepatitis C Virus Polymerase in Complex with Primer-Template RNA.” *Journal of Virology* 86, no. 12 (June 2012): 6503–11. doi:10.1128/JVI.00386-12.
87. Paul, David, Vanesa Madan, and Ralf Bartenschlager. “Hepatitis C Virus RNA Replication and Assembly: Living on the Fat of the Land.” *Cell Host & Microbe* 16, no. 5 (November 12, 2014): 569–79. doi:10.1016/j.chom.2014.10.008.
88. Vaillancourt, Frédéric H., Martine Brault, Louise Pilote, Nathalie Uyttersprot, Elias T. Gaillard, James H. Stoltz, Brian L. Knight, et al. “Evaluation of Phosphatidylinositol-4-Kinase III $\alpha$  as a Hepatitis C Virus Drug Target.” *Journal of Virology* 86, no. 21 (November 2012): 11595–607. doi:10.1128/JVI.01320-12.
89. Cameroni, Elisabetta, Claudio De Virgilio, and Olivier Deloche. “Phosphatidylinositol 4-Phosphate Is Required for Translation Initiation in *Saccharomyces Cerevisiae*.” *Journal of Biological Chemistry* 281, no. 50 (December 15, 2006): 38139–49. doi:10.1074/jbc.M601060200.
90. Reghellin, V., L. Donnici, S. Fenu, V. Berno, V. Calabrese, M. Pagani, S. Abrignani, F. Peri, R. De Francesco, and P. Neddermann. “NS5A Inhibitors Impair NS5A- PI4KIII $\alpha$  Complex Formation and Cause a Decrease of PI4P and Cholesterol Levels in HCV-Associated Membranes.” *Antimicrobial Agents and Chemotherapy*, September 15, 2014, AAC.03293–14. doi:10.1128/AAC.03293-14.

91. Hammond, Gerald R. V., and Tamas Balla. "Polyphosphoinositide Binding Domains: Key to Inositol Lipid Biology." *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*. Accessed March 9, 2015. doi:10.1016/j.bbalip.2015.02.013.
92. Bianco, Annalisa, Veronica Reghellin, Lorena Donnici, Simone Fenu, Reinaldo Alvarez, Chiara Baruffa, Francesco Peri, et al. "Metabolism of Phosphatidylinositol 4-Kinase III $\alpha$ -Dependent PI4P Is Subverted by HCV and Is Targeted by a 4-Anilino Quinazoline with Antiviral Activity." *PLoS Pathogens* 8, no. 3 (2012): e1002576. doi:10.1371/journal.ppat.1002576.
93. Madan, Vanesa, David Paul, Volker Lohmann, and Ralf Bartenschlager. "Inhibition of HCV Replication by Cyclophilin Antagonists Is Linked to Replication Fitness and Occurs by Inhibition of Membranous Web Formation." *Gastroenterology* 146, no. 5 (May 2014): 1361–72.e1–9. doi:10.1053/j.gastro.2014.01.055.
94. Chao, Ti-Chun, Wen-Chi Su, Jing-Ying Huang, Yung-Chia Chen, King-Song Jeng, Horng-Dar Wang, and Michael M. C. Lai. "Proline-Serine-Threonine Phosphatase-Interacting Protein 2 (PSTPIP2), a Host Membrane-Deforming Protein, Is Critical for Membranous Web Formation in Hepatitis C Virus Replication." *Journal of Virology* 86, no. 3 (February 2012): 1739–49. doi:10.1128/JVI.06001-11.
95. Park, Chul-Yong, Hyun-Jeong Jun, Takaji Wakita, Jae Hun Cheong, and Soon B. Hwang. "Hepatitis C Virus Nonstructural 4B Protein Modulates Sterol Regulatory Element-Binding Protein Signaling via the AKT Pathway." *The Journal of Biological Chemistry* 284, no. 14 (April 3, 2009): 9237–46. doi:10.1074/jbc.M808773200.
96. Wang, Hongliang, Jeffrey W. Perry, Adam S. Lauring, Petra Neddermann, Raffaele De Francesco, and Andrew W. Tai. "Oxysterol-Binding Protein Is a Phosphatidylinositol 4-Kinase Effector Required for HCV Replication Membrane Integrity and Cholesterol Trafficking." *Gastroenterology* 146, no. 5 (May 2014): 1373–85.e1–11. doi:10.1053/j.gastro.2014.02.002.
97. Mesmin, Bruno, Joëlle Bigay, Joachim Moser von Filseck, Sandra Lacas-Gervais, Guillaume Drin, and Bruno Antonny. "A Four-Step Cycle Driven by PI(4)P Hydrolysis Directs sterol/PI(4)P Exchange by the ER-Golgi Tether OSBP." *Cell* 155, no. 4 (November 7, 2013): 830–43. doi:10.1016/j.cell.2013.09.056.
98. Khan, Irfan, Divya S. Katikaneni, Qingxia Han, Lorena Sanchez-Felipe, Kentaro Hanada, Rebecca L. Ambrose, Jason M. Mackenzie, and Kouacou V. Konan. "Modulation of Hepatitis C Virus Genome Replication by Glycosphingolipids and Four-Phosphate Adaptor Protein 2." *Journal of Virology* 88, no. 21 (November 2014): 12276–95. doi:10.1128/JVI.00970-14.
99. Remmert, Michael, Andreas Biegert, Andreas Hauser, and Johannes Söding. "HHblits: Lightning-Fast Iterative Protein Sequence Searching by HMM-HMM Alignment." *Nature Methods* 9, no. 2 (February 2012): 173–75. doi:10.1038/nmeth.1818.
100. Söding, Johannes, Andreas Biegert, and Andrei N. Lupas. "The HHpred Interactive Server for Protein Homology Detection and Structure Prediction." *Nucleic Acids Research* 33, no. Web Server issue (July 1, 2005): W244–48. doi:10.1093/nar/gki408.
101. Finn, Robert D., Jaina Mistry, John Tate, Penny Coghill, Andreas Heger, Joanne E. Pollington, O. Luke Gavin, et al. "The Pfam Protein Families Database." *Nucleic Acids Research* 38, no. Database issue (January 2010): D211–22. doi:10.1093/nar/gkp985.
102. Yang, Jianyi, Renxiang Yan, Ambrish Roy, Dong Xu, Jonathan Poisson, and Yang Zhang. "The I-TASSER Suite: Protein Structure and Function Prediction." *Nature Methods* 12, no. 1 (January 2015): 7–8. doi:10.1038/nmeth.3213.
103. <http://zhanglab.ccmb.med.umich.edu/I-TASSER/about.html>
104. Suer, Silke, Albert Sickmann, Helmut E. Meyer, Friedrich W. Herberg, and Ludwig M. G. Heilmeyer. "Human Phosphatidylinositol 4-Kinase Isoform PI4K92." *European Journal of Biochemistry* 268, no. 7 (April 1, 2001): 2099–2106. doi:10.1046/j.1432-1327.2001.02089.x.

105. Nolen, Brad, Susan Taylor, and Gourisankar Ghosh. "Regulation of Protein Kinases; Controlling Activity through Activation Segment Conformation." *Molecular Cell* 15, no. 5 (September 10, 2004): 661–75. doi:10.1016/j.molcel.2004.08.024.
106. Sievers, Fabian, Andreas Wilm, David Dineen, Toby J. Gibson, Kevin Karplus, Weizhong Li, Rodrigo Lopez, et al. "Fast, Scalable Generation of High-Quality Protein Multiple Sequence Alignments Using Clustal Omega." *Molecular Systems Biology* 7 (2011): 539. doi:10.1038/msb.2011.75.
107. Waterhouse, Andrew M., James B. Procter, David M. A. Martin, Michèle Clamp, and Geoffrey J. Barton. "Jalview Version 2—a Multiple Sequence Alignment Editor and Analysis Workbench." *Bioinformatics* 25, no. 9 (May 1, 2009): 1189–91. doi:10.1093/bioinformatics/btp033
108. Kelley, Lawrence A., Stefans Mezulis, Christopher M. Yates, Mark N. Wass, and Michael J. E. Sternberg. "The Phyre2 Web Portal for Protein Modeling, Prediction and Analysis." *Nature Protocols* 10, no. 6 (June 2015): 845–58. doi:10.1038/nprot.2015.053.
109. Zhang, Yang, and Jeffrey Skolnick. "TM-Align: A Protein Structure Alignment Algorithm Based on the TM-Score." *Nucleic Acids Research* 33, no. 7 (2005): 2302–9. doi:10.1093/nar/gki524.
110. Mitchell, Alex, Hsin-Yu Chang, Louise Daugherty, Matthew Fraser, Sarah Hunter, Rodrigo Lopez, Craig McAnulla, et al. "The InterPro Protein Families Database: The Classification Resource after 15 Years." *Nucleic Acids Research* 43, no. Database issue (January 2015): D213–21. doi:10.1093/nar/gku1243.
111. Bian, Yangyang, Chunxia Song, Kai Cheng, Mingming Dong, Fangjun Wang, Junfeng Huang, Deguang Sun, Liming Wang, Mingliang Ye, and Hanfa Zou. "An Enzyme Assisted RP-RPLC Approach for in-Depth Analysis of Human Liver Phosphoproteome." *Journal of Proteomics* 96 (January 16, 2014): 253–62. doi:10.1016/j.jprot.2013.11.014.
112. Ashkenazy, Haim, Elana Erez, Eric Martz, Tal Pupko, and Nir Ben-Tal. "ConSurf 2010: Calculating Evolutionary Conservation in Sequence and Structure of Proteins and Nucleic Acids." *Nucleic Acids Research* 38, no. Web Server issue (July 2010): W529–33. doi:10.1093/nar/gkq399.
113. Tusnády, G. E., and I. Simon. "Principles Governing Amino Acid Composition of Integral Membrane Proteins: Application to Topology Prediction." *Journal of Molecular Biology* 283, no. 2 (October 23, 1998): 489–506. doi:10.1006/jmbi.1998.2107.
114. Käll, Lukas, Anders Krogh, and Erik L. L. Sonnhammer. "Advantages of Combined Transmembrane Topology and Signal Peptide Prediction—the Phobius Web Server." *Nucleic Acids Research* 35, no. Web Server issue (July 2007): W429–32. doi:10.1093/nar/gkm256.
115. Mayya, Viveka, Deborah H. Lundgren, Sun-Il Hwang, Karim Rezaul, Linfeng Wu, Jimmy K. Eng, Vladimir Rodionov, and David K. Han. "Quantitative Phosphoproteomic Analysis of T Cell Receptor Signaling Reveals System-Wide Modulation of Protein-Protein Interactions." *Science Signaling* 2, no. 84 (2009): ra46. doi:10.1126/scisignal.2000007.
116. Studier, F. W., and B. A. Moffatt. "Use of Bacteriophage T7 RNA Polymerase to Direct Selective High-Level Expression of Cloned Genes." *Journal of Molecular Biology* 189, no. 1 (May 5, 1986): 113–30.
117. Miroux, B., and J. E. Walker. "Over-Production of Proteins in Escherichia Coli: Mutant Hosts That Allow Synthesis of Some Membrane Proteins and Globular Proteins at High Levels." *Journal of Molecular Biology* 260, no. 3 (July 19, 1996): 289–98. doi:10.1006/jmbi.1996.0399.
118. Grunberg-Manago, M. "Messenger RNA Stability and Its Role in Control of Gene Expression in Bacteria and Phages." *Annual Review of Genetics* 33 (1999): 193–227. doi:10.1146/annurev.genet.33.1.193.
119. Ikemura, T. "Correlation between the Abundance of Escherichia Coli Transfer RNAs and the Occurrence of the Respective Codons in Its Protein Genes." *Journal of Molecular Biology* 146, no. 1 (February 15, 1981): 1–21.
120. Thomas, J. G., A. Ayling, and F. Baneyx. "Molecular Chaperones, Folding Catalysts, and the Recovery of Active Recombinant Proteins from E. Coli. To Fold or to Refold." *Applied Biochemistry and Biotechnology* 66, no. 3 (June 1997): 197–238.

121. Diaporama du cours Hercules 2011 « Solution X-ray Scattering from Biological Macromolecules» de Javier Pérez
122. Petoukhov, Maxim V., Daniel Franke, Alexander V. Shkumatov, Giancarlo Tria, Alexey G. Kikhney, Michal Gajda, Christian Gorba, Haydyn D. T. Mertens, Petr V. Konarev, and Dmitri I. Svergun. "New Developments in the ATSAS Program Package for Small-Angle Scattering Data Analysis." *Journal of Applied Crystallography* 45, no. Pt 2 (April 1, 2012): 342–50. doi:10.1107/S0021889812007662.
123. <http://www.embl-hamburg.de/biosaxs.html>
124. P. V. Konarev et al. *PRIMUS*: a Windows PC-based system for small-angle scattering data analysis. *Journal of Applied Crystallography* . 2003
125. Fanchon, Eric. *Structure and Dynamics of Biomolecules: Neutron and Synchrotron Radiation for Condensed Matter Studies*. Oxford ; New York: OUP Oxford, 2000.
126. Mertens, Haydyn D. T., and Dmitri I. Svergun. "Structural Characterization of Proteins and Complexes Using Small-Angle X-Ray Solution Scattering." *Journal of Structural Biology* 172, no. 1 (October 2010): 128–41. doi:10.1016/j.jsb.2010.06.012.
127. Bernadó, Pau, and Martin Blackledge. "A Self-Consistent Description of the Conformational Behavior of Chemically Denatured Proteins from NMR and Small Angle Scattering." *Biophysical Journal* 97, no. 10 (November 15, 2009): 2839–45. doi:10.1016/j.bpj.2009.08.044.
128. Svergun, D I. "Restoring Low Resolution Structure of Biological Macromolecules from Solution Scattering Using Simulated Annealing." *Biophysical Journal* 76, no. 6 (June 1999): 2879–86.
129. Svergun, D I, M V Petoukhov, and M H Koch. "Determination of Domain Structure of Proteins from X-Ray Solution Scattering." *Biophysical Journal* 80, no. 6 (June 2001): 2946–53.
130. M.Kozin & D.Svergun (2001) Automated matching of high- and low-resolution structural models *J Appl Cryst.* 34, 33-41.
131. V. V. Volkov and D. I. Svergun (2003). Uniqueness of ab-initio shape determination in small-angle scattering.*J. Appl. Cryst.* 36, 860-864.
132. Main, Ewan R. G., Alan R. Lowe, Simon G. J. Mochrie, Sophie E. Jackson, and Lynne Regan. "A Recurring Theme in Protein Engineering: The Design, Stability and Folding of Repeat Proteins." *Current Opinion in Structural Biology* 15, no. 4 (August 2005): 464–71. doi:10.1016/j.sbi.2005.07.003.
133. Cortajarena, Aitziber L., and Lynne Regan. "Calorimetric Study of a Series of Designed Repeat Proteins: Modular Structure and Modular Folding." *Protein Science: A Publication of the Protein Society* 20, no. 2 (February 2011): 336–40. doi:10.1002/pro.564.
134. Moraes, Isabel, Gwyndaf Evans, Juan Sanchez-Weatherby, Simon Newstead, and Patrick D. Shaw Stewart. "Membrane Protein Structure Determination — The next Generation." *Biochimica et Biophysica Acta* 1838, no. 1 (January 2014): 78–87. doi:10.1016/j.bbamem.2013.07.010.
135. Perez-Iratxeta, Carolina, and Miguel A. Andrade-Navarro. "K2D2: Estimation of Protein Secondary Structure from Circular Dichroism Spectra." *BMC Structural Biology* 8 (2008): 25. doi:10.1186/1472-6807-8-25.
136. Love, Robert A., Oleg Brodsky, Michael J. Hickey, Peter A. Wells, and Ciarán N. Cronin. "Crystal Structure of a Novel Dimeric Form of NS5A Domain I Protein from Hepatitis C Virus." *Journal of Virology* 83, no. 9 (May 2009): 4395–4403. doi:10.1128/JVI.02352-08.
137. Lim, Precious J., Udayan Chatterji, Daniel Cordek, Suresh D. Sharma, Jose A. Garcia-Rivera, Craig E. Cameron, Kai Lin, Paul Targett-Adams, and Philippe A. Gally. "Correlation between NS5A Dimerization and Hepatitis C Virus Replication." *Journal of Biological Chemistry* 287, no. 36 (August 31, 2012): 30861–73. doi:10.1074/jbc.M112.376822.

138. Anderson, D., Harris, R., Polayes, D., Ciccarone, V., Donahue, R., Gerard, G., and Jessee, J. (1996) Rapid Generation of Recombinant Baculoviruses and Expression of Foreign Genes Using the Bac-ToBac® Baculovirus Expression System. *Focus* 17, 53-58
139. Yang, Haijuan, Derek G. Rudge, Joseph D. Koos, Bhamini Vaidialingam, Hyo J. Yang, and Nikola P. Pavletich. "mTOR Kinase Structure, Mechanism and Regulation." *Nature* 497, no. 7448 (May 9, 2013): 217–23. doi:10.1038/nature12122.
140. Sibanda, Bancinyane L., Dimitri Y. Chirgadze, and Tom L. Blundell. "Crystal Structure of DNA-PKcs Reveals a Large Open-Ring Cradle Comprised of HEAT Repeats." *Nature* 463, no. 7277 (January 7, 2010): 118–21. doi:10.1038/nature08648.
141. Schwede, Torsten. "Protein Modeling: What Happened to the 'Protein Structure Gap'?" *Structure (London, England: 1993)* 21, no. 9 (September 3, 2013): 1531–40. doi:10.1016/j.str.2013.08.007.
142. Zhang, Yang. "Protein Structure Prediction: Is It Useful?" *Current Opinion in Structural Biology* 19, no. 2 (April 2009): 145–55. doi:10.1016/j.sbi.2009.02.005.
143. Dujardin, Marie, Vanesa Madan, Roland Montserret, Puneet Ahuja, Isabelle Huvent, Helene Launay, Arnaud Leroy, et al. "A Proline-Tryptophan Turn in the Intrinsically Disordered Domain 2 of NS5A Protein Is Essential for Hepatitis C Virus RNA Replication." *The Journal of Biological Chemistry* 290, no. 31 (July 31, 2015): 19104–20. doi:10.1074/jbc.M115.644419.